

---

# Challenges in Speech Recognition and Translation of High-Value Low-Density Polysynthetic Languages

**Judith L. Klavans**  
**John Morgan**  
**Stephen LaRocca**  
**Jeffrey Micher**  
**Clare Voss**

judith.l.klavans.civ@mail.mil  
john.j.morgan50.civ@mail.mil  
stephen.a.larocca.civ@mail.mil  
jeffrey.c.micher.civ@mail.mil  
clare.r.voss.civ@mail.mil

Multilingual Computing and Analytics Branch, Army Research Laboratory,  
Adelphi, Maryland 20783, USA

---

## Abstract

The focus of this paper is on setting out a framework for experiments on using the latest machine learning techniques over speech and text data collections of highly complex languages. We are in the process of creating comparable and consistent databases with associated processing technologies of some of the world's most challenging languages, polysynthetic languages, i.e. those where one long word can express the meaning contained in a multi-word sentence in languages like English. We present an end-to-end system for Automatic Speech Recognition (ASR) and Machine Translation (MT) involving Artificial Intelligence approaches of machine learning (ML). The ML framework uses deep learning since the networks we are sharing are deep in nature; this deep variant of Multi-Task ML (MTML) embodies human-like AI abilities to learn a language with small amounts of input thereby achieving a degree of AI. We explore recurrent neural networks (RNNs), long and short term memory network (L-STMs), bidirectional LSTMs (BiLSTM) and convolutional NNs (CNN) to compare and evaluate results.

## 1. Motivation

The government and military have to respond to and communicate in languages that present themselves in the field – whether for humanitarian aid, intelligence or other operational requirements. Currently, the government and military have many language requirements, ranging from interacting with coalition forces to public affairs to on-the-ground soldier interaction with foreign citizens to intelligence. To quote a current Program Manager at DARPA<sup>1</sup> in the Information Innovation Office (I2O):

*“We do not know what language will be next in line for military and national defense needs. Thus, we need to be prepared with technology to handle any language of any complexity, and we need the capability to ramp up with small amounts of data.”*

---

<sup>1</sup> Dr. Boyan Onyshkevych, personal communication.

<sup>2</sup> <https://www.nytimes.com/2017/10/04/world/africa/special-forces-killed-niger.html>

Increased globalization has led to an urgent need for even more and varied language capabilities than in the past. As Army Chief of Staff, Gen. Mark Milley said in response to the gap in intelligence leading to the recent ambush against US troops in Niger:

*"We are training, advising and assisting indigenous armies all over the world. And I anticipate and expect that will increase not decrease in years to come,"<sup>2</sup>*

This paper presents strategies for addressing the computational and linguistic challenges posed by such complex languages. We address specifically the areas of automatic speech recognition and MT research and development in government and military settings.

## 2. Research Goals

The focus of this paper is on setting out a framework for experiments on using the latest machine learning techniques over speech, text, and data collections of highly complex languages. We are in the process of creating comparable databases with associated processing technologies of some of the world's most challenging languages, those where one long word can express the meaning contained in a multi-word sentence in languages like English. These are called polysynthetic languages. To illustrate, consider the following example from Inuktitut, one of the official languages of the Territory of Nunavut in Canada. The morpheme **-tusaa-** (shown in boldface below) is the root, and all the other morphemes are synthetically combined with it in one unit.<sup>3</sup>

- (1) **tusaa**-tsia-runna-nngit-tu-alu-u-junga  
hear-well-be.able-NEG-DOER-very-BE-PART.1.S  
'I can't hear very well.'

Kabardian (Circassian), from the Northwest Caucasus, also shows this phenomenon, with the root **-še-** shown in boldface below:

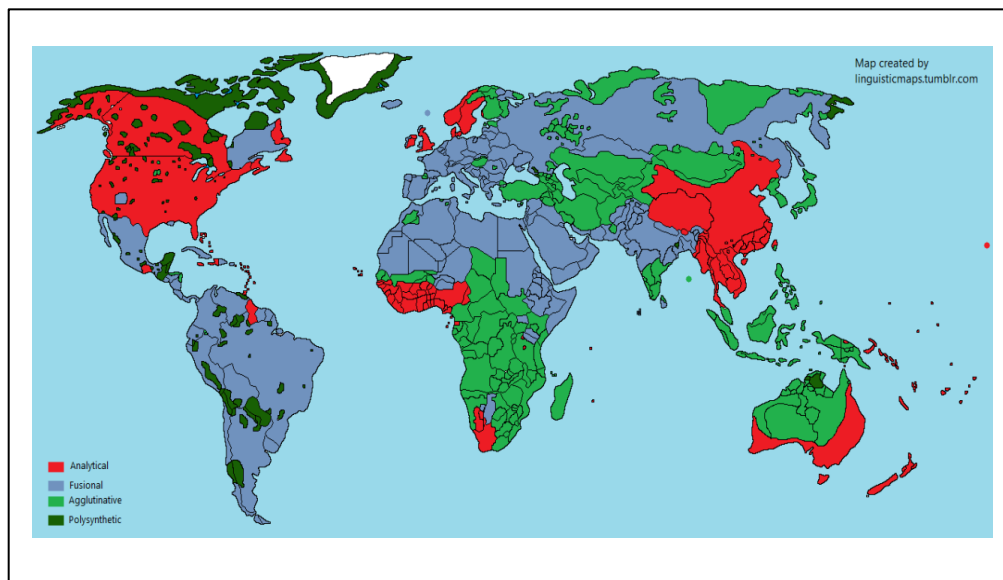
- (2) wə-q'ə-d-ej-z-ye-še-ž'e-f-a-te-q'əm  
2SG.OBJ-DIR-LOC-3SG.OBJ-1SG.SUBJ-CAUS-lead-COMPL-POTENTIAL-PAST-PRF-NEG  
'I would not let you bring him right back here.'

Polysynthetic languages are spoken all over the globe and are richly represented among Native North and South American families. Many polysynthetic languages are among the world's most endangered languages,<sup>4</sup> with fragmented dialects and communities struggling to preserve their linguistic heritage. In particular, polysynthetic languages can be found in the US Southwest (Southern Tiwa, Kiowa Tanoan family), Canada, Mexico (Nahuatl, Uto-Aztecan family), and Central Chile (Mapudungun, Araucanian), as well as in Australia (Nunggubuyu, Macro-Gunwinyguan family), Northeastern Siberia (Chukchi and Koryak, both from the Chukotko-Kamchatkan family), and India (Sora, Munda family), as shown in the map below (Figure 1).

<sup>2</sup> <https://www.nytimes.com/2017/10/04/world/africa/special-forces-killed-niger.html>

<sup>3</sup> Abbreviations follow the Leipzig Glossing Rules; additional glosses are spelled out in full.

<sup>4</sup> In fact, the majority of the languages spoken in the world today are endangered and disappearing fast (See Bird, 2009). Estimates are that, of the approximately 7000 languages in the world today, at least one disappears every day (<https://www.ethnologue.com>).



**Figure 1: Polysynthetic Languages<sup>5</sup>**

Although there are many definitions of polysynthesis, there is often confusion on what constitutes the exact criteria and phenomena (Mithun 2017). Even authoritative sources categorize languages in conflicting ways.<sup>6</sup> Typically, polysynthetic languages demonstrate holophrasis, i.e. the ability of an entire sentence to be expressed in what is considered by native speakers to be just one word (Bird 2009). In linguistic typology, the opposite of polysynthesis is *isolation*. Polysynthesis technically (etymologically) refers to how many morphemes there are per word. Using that criterion, the typological continuum can be represented as follows:

(3) isolating/analytic languages > synthetic languages > polysynthetic languages

Adding another dimension of morphological categorization, languages can be distinguished by the degree of clarity of morpheme boundaries. If we apply this criterion, languages can be categorized according to the following typological continuum:

(4) agglutinating > mildly fusional > fusional

Thus, a language might be characterized overall as polysynthetic and agglutinating, that is, generally a high number of morphemes per word, with clear boundaries between morphemes and thus easily segmentable. Another language might be characterized as polysynthetic and fusional, so again, many morphemes per word, but so many phonological and other processes have occurred that segmenting morphemes becomes more challenging.

<sup>5</sup> <http://linguisticmaps.tumblr.com/post/120857875008/513-morphological-typology-tonal-languages>. Map by Rodrigo Pereira.

<sup>6</sup> For example, the article in the *Oxford Research Encyclopedia of Linguistics* on “Polysynthesis: A Diachronic and Typological Perspective” by Michael Fortescue (Fortescue, 2016), a well-known expert on polysynthesis, lists Aymara as possibly polysynthetic, whereas others designate it as agglutinative (<http://www.native-languages.org>).

So far, we have discussed the morphological aspects of polysynthesis. Polysynthesis also has a number of syntactic ramifications, richly explored in the work of Baker (Baker 1997; 2002). He proposes a cluster of correlated syntactic properties associated with polysynthesis. Here we will mention just two of these properties: rich agreement (with the subject, direct object, indirect object, and applied objects if present) and omission of free-standing arguments (pro-drop).

Polysynthetic languages are of interest for both theoretical and practical reasons, as discussed more fully in the next section. On the theoretical side, these languages offer a potentially unique window into human cognition and language capabilities as well as into language acquisition (Mithun 1989; Greenberg 1960; Comrie 1981; Fortescue 1994; Fortescue et al. 2017). On the practical side, they offer significant obstacles to accurate linguistic analysis as well as to computational modeling.

### 3. Some Computational Challenges of Polysynthetic Languages

Polysynthetic languages pose unique challenges for traditional computational systems (Byrd et al. 1986). Even in allegedly cross-linguistic or typological analyses of specific phenomena, e.g. in forming a theory of clitics and cliticization (Klavans 1995), finding the full range of language types on which to test hypotheses proves difficult. Often, the data is simply not available so claims cannot neither refuted nor supported fully.

One of the underlying causes of this difficulty is that there are many languages for which a clear lexical division between nouns and verbs has been challenged; these languages are characterized by a large class of roots that are used either nominally or verbally, and many of these languages typically have polysynthetic features (cf. Lois & Vapnarsky 2006 for Amerindian, Aranovich 2013 for Austronesian, Testelefs et al. 2009 for Adyghe, Davis & Matthewson 2009, Watanabe 2017 for Salish). Without a clear definition of what counts as a verb and what counts as a noun, there is no reliable way to compute significant correlations. Thus, a deeper understanding of polysynthetic phenomena may well contribute to a more nuanced understanding of cross-language comparisons and generalizations and enable researchers to pose meaningful and answerable questions about comparative features across languages.

On the practical side, many morphologically complex languages are crucial to purposes ranging from health care,<sup>7</sup> search and rescue, to the maintenance of cultural history (Fortescue et al. 2017). Add to this the interest in low-resource languages (from Inuktitut and Yup'ik in the North and East of Canada with over 35,000 speakers, and all the way to Northwest Caucasian), which is important for linguistic, cultural and governmental reasons. Many of the data collections in these languages, when annotated and aligned well, can serve as input to systems to automatically create correspondences, and these in turn can be useful to teachers in creating resources for their learners (Adams, Neubig, Cohn, & Bird 2015). These languages are generally not of immediate commercial value, and yet the research community needs to cope with

---

<sup>7</sup> For example, the USAID has funded a program in the mountains of Ecuador to provide maternal care in Quechua-dominant areas to reduce maternal and infant mortality rates, taking into account local cultural and language needs (<https://www.usaidassist.org>). Quechua is highly agglutinative, not polysynthetic; it is spoken by millions of speakers and has few corpora with limited annotation.

fundamental issues of language complexity. Finally, many of these understudied languages occur in areas that are key for health concerns (e.g. the AIDS epidemic) and international security. Consequently, research on these languages could have unanticipated benefits on many levels.

Recent research (e.g. Micher 2016) has applied neural nets to one polysynthetic language towards creating a feasible model for machine translation. As for speech recognition, longer words are generally less prone to error (Shinozaki & Furui 2001); this accounts for the fact that under 70% word accuracy is useful for keyword spotting, as shown in the IARPA Babel project<sup>8</sup>. On the other hand, if a language has only very long “words” encompassing all the nouns, verbs, clitics, affixes and particles, then these languages might not conform to established principles. At the same time, morphological and syntactic processing of polysynthetic languages pose specific challenges due to the blur between the more usual morphology-syntax distinction (Baker 1996). On low-resource language speech recognition, based on our experience with a range of language types, we hypothesize that the most effective units of recognition might be morphemes, although many of these morphemes might have a variety of possible surface forms. Because of the sentential nature of words in these languages, they can constitute a number of unique forms, raising intriguing speech recognition challenges.

#### 4. Ongoing Language Research at the Army Research Laboratory

This paper provides an overview of one aspect of multilingual language research at the Army Research Laboratory, presenting the approaches used in polysynthetic languages. Figure 2 below shows which aspects of the project are being addressed. In the presentation, we will discuss technical details of each component and discuss further the novel methodological contributions of the research.

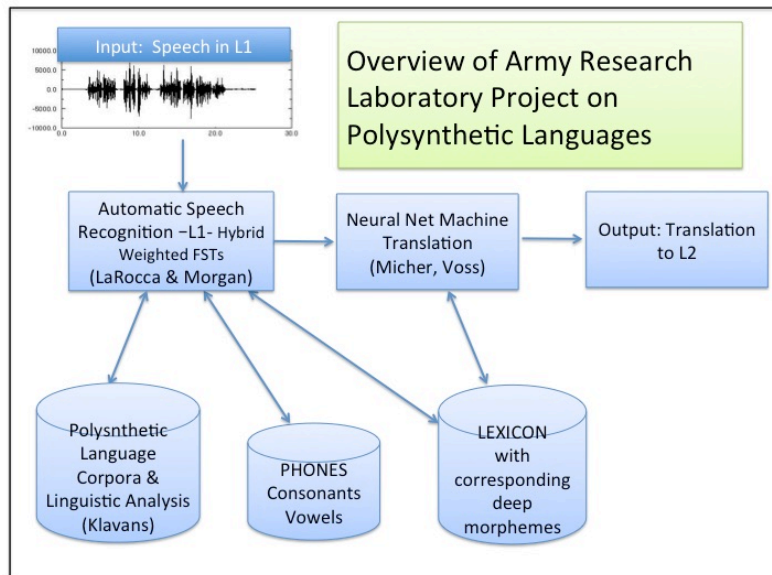


Figure 2: Overview of Speech-MT Polysynthetic Language Architecture

<sup>8</sup> <https://www.gov/index.php/research-programs/babel>

ARL has demonstrated leading technologies in the field with critical expertise. We are planning on developing systems, capable of performing speech translation. We are applying machine learning techniques using neural network approaches e.g. segmental recurrent neural networks (Kong et al. 2015, Micher 2017) and byte-pair encoding Sennrich, Haddow, Birch 2015) to several challenging problems for polysynthetic language analysis and processing. For ASR, we have implemented adaptive learning for iterative ASR, incorporating principles from the Kaldi toolkit<sup>9</sup> with modifications as required by different workflows and tasks.

## 5. Corpus Collection - Electronically-available resources

Only recently have researchers started collecting well-designed corpora for polysynthetic languages, e.g. for Circassian (Arkhangelskiy & Lander 2016) or Arapaho (Kazeminejad et al. 2017). There is an urgent need for documentation, archiving, creation of corpora and teaching materials that are specific to polysynthetic languages. Documentation and corpus-building challenges arise for many languages, but the complex morphological makeup of polysynthetic languages makes consistent documentation particularly difficult.

The more language data that is gathered and accurately analyzed, the deeper cross-linguistic analyses can be conducted which in turn will contribute to a range of fields including linguistic theory, language teaching and lexicography. For example, in examining cross-linguistic analyses of headedness, Polinsky (2012) gathered data to examine the question of whether the noun-verb ratio differs across headedness types across a wide sample of language types. However, she notes that:

*“[T]he seemingly simple question of counting nouns and verbs is a quite difficult one; even obtaining data about the overall number of nouns and verbs proves to be an immense challenge. The ultimate consequence is that linguists lack reasonable tools to compare languages with respect to their lexical category size. Cooperation between theoreticians and lexicographers is of critical importance: just as comparative syntax received a big boost from the micro-comparative work on closely related languages (Romance; Germanic; Semitic), so micro-comparative WordNet building may lead to important breakthroughs that will benefit the field as a whole.” (Polinsky, 2012, p. 351)*

In recent years, there has been a surge of major research on many of these languages. For example, the first Endangered Languages (ELs) Workshop held in conjunction with ACL was held in 2014 and the second in 2017.<sup>10</sup> The National Science Foundation and the National Endowment for the Humanities jointly fund a program for research on ELs.<sup>11</sup> The US government through IARPA and DARPA both have programs for translation, including for low resource languages.<sup>12</sup> The IARPA BABEL project focused on keyword search over speech for a variety of typologically different languages, including some with polysynthetic features.

---

<sup>9</sup> <http://kaldi-asr.org/doc/pages.html>

<sup>10</sup> <http://www.acsu.buffalo.edu/~jcgood/ComputEL.html>;  
<http://altlab.artsrn.ualberta.ca/computel-2/>.

<sup>11</sup> [https://www.nsf.gov/funding/pgm\\_summ.jsp?pims\\_id=12816](https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=12816);  
<https://www.neh.gov/grants/manage/general-information-neh-nsf-documenting-endangered-languages-fellowships>.

<sup>12</sup> MATERIAL, <https://www.iarpa.gov/index.php/research-programs/material> and LORELEI, <https://www.darpa.mil/program/low-resource-languages-for-emergent-incidents>, respectively.

Concomitant with the collection and cataloging of corpora, we are working with colleagues especially from the the NSF-funded EL-STEC Shared Task Evaluation Campaign project<sup>13</sup> on a future *shared task* in order to bring linguists and computational linguists together around the common area: accuracy in data analysis. We aim to formulate a shared task that meets the goals outlined in Levow, et al. (2017), namely, to “align the interests of the speech and language processing communities with those of ... language documentation communities...”, guided by their design principles of realism, typological diversity, accessibility of the shared task, accessibility of the resulting software, extensibility and nuanced evaluation.

## 6. Future Research and Applications

Our next steps involve a two-phase approach, one on the ASR input and then one on the MT side (as shown in Figure 2.) On the ASR side, we plan to use Multi-Task Learning (MTL) (Caruana 1997), using corpora from multiple languages. Multitask Learning (also known as Multi-Task Machine Learning MTML) is an approach to inductive transfer that improves generalization by using the domain information contained in the training signals of related tasks as an inductive bias. It does this by learning tasks in parallel while using a shared representation; what is learned for each task can help other tasks be learned better. MTL is an established machine learning framework that has been applied to multiple domains. The ASR problem, however, brings specific language problems to any machine learning approach. As noted in Hasegawa-Johnson 2017:

*To date, ASR has failed to achieve its potential, because successful ASR requires very large labeled corpora; the human transcribers must be computer-literate, and they must be native speakers of the language being transcribed. Large corpora are beyond the resources of most under-resourced language communities; we have found that transcribing even one hour of speech may be beyond the reach of communities that lack large-scale government funding. (Hasegawa-Johnson et al. 2017, p. 50)*

This deep variant of MTML that we use embodies human-like AI abilities to learn a language with small amounts of input thereby achieving a degree of AI. We build on related techniques, widely used in the ASR community (Povey et al. 2011). The original contribution consists of using a range of conversational modalities (news, dialog, read speech) as sources of data in order to realize the potential for dissimilar input to contribute to more robust output. We hypothesize that the MTL technique can capture features characteristic of the target Low Resource language across dissimilar modalities and similar languages. Our approach is reported in LaRocca and Morgan 2018, to appear.

On the Machine Translation side, the research questions to be addressed in future work include methods to improve the performance of the existing Uqailaut morphological analyzer for the Inuktitut (Farley, 2009) making use of a variety of neural network approaches; improvements over a baseline statistical machine translation (SMT) English-Inuktitut system by using alternate subword units with a neural network architecture; diagnosis of which subword units yield the most improvement; determining how a pipelined English-Inuktitut translation

---

<sup>13</sup> <http://depts.washington.edu/uwcl/el-stec/index.php>

system, with deep morpheme translation plus deep-to-surface sequence-to-sequence model performs compared with the best subword system; and then exploring the use of hierarchical structures over morphemes in a novel approach to improve over the best subword system.

From an applications perspective, the outcomes of the research will be useful for a wide range of applications including collaboration with coalition forces and civil affairs requirements, in particular. From a theoretical perspective, we contribute to a deeper understanding of the effectiveness of neural network architectures which take context into consideration, for example, a recurrent neural network (RNN), a long- short term memory network (LSTM), a bidirectional LSTM (BiLSTM), or a convolutional neural network (CNN). We will reveal necessary modifications in order for successful low-resource ASR and MT. Finally, from the perspective of language revitalization and contributions to native communities, we explore tools that could be useful to teachers and language analysts as we reach the future goal of enabling a deep understanding of language across types and both their superficial and underlying features.

To conclude, we have set out a strategy and approach for an end-to-end speech recognition system along with machine translation that involves developing novel machine learning techniques and computational approaches for low-resource polysynthetic languages.

**Acknowledgements:** Thanks to Roland Kuhn and Anna Kazantseva, National Research Council of Canada who are part of the organizing committee for the upcoming workshop on Polysynthetic Languages to be held at COLING 2018, in Santa Fe, New Mexico on August 25, 2018. This workshop will be the first where both researchers and practitioners working on polysynthetic languages will discuss common problems and difficulties, and it is intended as the capstone to establishing possible collaborations and ongoing partnerships. We acknowledge Lori Levin, Carnegie Mellon University, for many helpful discussions on the nature of polysynthesis, and we are grateful to Maria Polinsky, University of Maryland, and Omer Preminger, University of Maryland, for enabling us to dig even deeper into some of the more fine-grained aspects of polysynthesis. All errors are, of course, our own.

## References

- Adams, O., Neubig, G., Cohn, T., & Bird, S. (2015). Inducing bilingual lexicons from small quantities of sentence-aligned phonemic transcriptions. *Proceedings of the International Workshop on Spoken Language Translation (IWSLT 2015)*. Da Nang, Vietnam.
- Aranovich, R. (2013). Transitivity and polysynthesis in Fijian. *Language* 89: 465-500.
- Arkhangelskiy, T. A., & Lander, Y. A. (2016). Developing a polysynthetic language corpus: problems and solutions. *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2016"*, June 104, 2016.



- Baker, M. C. (1996). *The polysynthesis parameter*. New York: Oxford University Press.
- Baker, M.C. (2002). *Atoms of language*. New York: Basic Books.
- Bird, S. (2009). Natural language processing and linguistic fieldwork. *Computational Linguistics*, 35 (3), 469-474.
- Byrd, R. J., Klavans, J. L., Aronoff, M., & Anshen, F. (1986). Computer methods for morphological analysis. *Proceedings of the 24th annual meeting on Association for Computational Linguistics* (pp. 120-127). Stroudsburg, PA. Association for Computational Linguistics.
- Caruana, Rich, (1997) "Multitask Learning." *Machine Learning*, Vol. 28, pp. 41-75, Kluwer Academic Publishers.
- Comrie, B. (1981). *Language Universals and Linguistic Typology*. Oxford: Blackwell.
- Davis, H., & Mattewson, L. (2009). Issues in Salish syntax and semantics. *Language and Linguistics Compass* 3, 1097-1166.
- Farley, B. (2009). *The Uqailaut Project*. Retrieved from Inuktitut Computing: <http://www.inuktitutcomputing.ca/Uqailaut/info.php>
- Fortescue, M. (1994). Polysynthetic morphology. (R. E. al., Ed.) *The encyclopedia of language and linguistic*, 5, 2600–2602.
- Fortescue, M. (2016). Polysynthesis: A Diachronic and Typological Perspective. In M. Aronoff (ed.) *Oxford Encyclopedia of Linguistics*. Oxford, Oxford, England: Oxford University Press.
- Fortescue, M., Mithun, M., & Evans, N. (Eds.). (2017). *The Oxford Handbook of Polysynthesis*. Oxford: Oxford University Press.
- Greenberg, J. H. (1960). A quantitative approach to the morphological typology of language. *International Journal of Linguistics*, 26, 178–194.
- Kazeminejad, G., Cowell, A., & Hulden, M. (2017). Creating lexical resources for polysynthetic languages—the case of Arapaho. *Proceedings of the 2nd Workshop on the Use of Computational Methods in the Study of Endangered Languages* (pp. 10-18). Honolulu: Association for Computational Linguistics.
- Klavans, J. L. (1995). *On Clitics and Cliticization: The Interaction of Morphology, Phonology, and Syntax*. New York: Garland.

- Kong, L., Dyer, C., & Smith, N. (2015). Segmental Recurrent Neural Networks. *CoRR*. Retrieved from <http://arxiv.org/abs/1511.06018>.
- LaRocca, Stephen and John Morgan (2018, to appear) “Incorporating MT into a Bi-directional Speech Translation System for U.S. Army units”, Paper to be presented at the Association for Machine Translation in the Americas conference (AMTA 2018), Boston, Massachusetts. March 17-21, 2018.
- Levow, G.-A., Bender, E., Littell, P., Howell, K., Chelliah, S., Crowgey, J., et al. (2017). STREAMLInED Challenges: Aligning Research Interests with Shared Tasks. *Proceedings of ComputEL-2: 2nd Workshop on Computational Methods for Endangered Languages*.
- Lois, X., & Vapnarsky, V. (2006.). Root indeterminacy and polyvalence in Yucatecan Mayan languages. In X. Lois, & V. Vapnarsky (Eds.). *Lexical categories and root clauses in Amerindian languages* (pp. 69-115). Bern: Peter Lang.
- Micher, Jeffrey (2016) “Machine Translation for a Low-Resource, Polysynthetic Language” Presentation at AMTA 2016. Austin, Texas.
- Micher, J. (2017). Improving Coverage of an Inuktitut Morphological Analyzer Using a Segmental Recurrent Neural Network. *Proceedings of the 2nd Workshop on the Use of Computational Methods in the Study of Endangered Languages* (pp. 101-106). Honolulu, HI: Association for Computational Linguistics.
- Mithun, M. (1989). The acquisition of polysynthesis. *Journal of Child Language*, 16, 285–312.
- Mithun, M. (2017). Argument marking in the polysynthetic verb and its implications. In M. Fortescue, M. Mithun, & N. Evans (Eds.), *The Oxford Handbook of Polysynthesis* (pp. 30-58). Oxford, UK: Oxford University Press.
- Polinsky, M. (2012). Headedness, again. *UCLA Working Papers in Linguistics, Theories of Everything*. 17, pp. 348-359. Los Angeles: UCLA.
- Povey, Daniel & Ghoshal, Arnab & Boulianne, Gilles & Burget, Lukáš & Glembek, Ondrej & Goel, Nagendra & Hannemann, Mirko & Motlíček, Petr & Qian, Yanmin & Schwarz, Petr & Silovský, Jan & Stemmer, Georg & Vesel, Karel. (2011). The Kaldi speech recognition toolkit. IEEE 2011 Workshop on Automatic Speech Recognition and Understanding.
- Sennrich, R., Haddow, B., & Birch, A. (2015). Neural Machine Translation of Rare Words with Subword Units. *CoRR*, *abs/1508.07909*. Retrieved from <http://arxiv.org/abs/1508.07909>

Shinozaki, T., Furui, S., 2001. Error analysis using decision trees in spontaneous presentation speech recognition. In: Proceedings of the Automatic Speech Recognition and Understanding Conference. Trento, Italy.

Testelets Ya. (ed.). (2009). *Aspekty polisintetizma: Očerki po grammatike adyghejskogo jazyka [Aspects of polysynthesis: Essays on Adyghe grammar]*, (pp. 17-120). Moscow: Russian University for the Humanities.

Watanabe, H. (2017). The polysynthetic nature of Salish. In Fortescue, M., Mithun, M., & Evans, N. (Eds.). (2017). *The Oxford Handbook of Polysynthesis* (pp. 623-642). Oxford: Oxford University Press.