Expanding n-gram analytics in ELAN and a case study for sign synthesis

Rosalee Wolfe[1], John C. McDonald[1], Larwan Berke[2], Marie Stumbo[1]

[1]DePaul University, Chicago, IL [2]Gallaudet University, Washington, DC
{wolfe, jmcdonald}@cs.depaul.edu, mstumbo@mail.depaul.edu, larwan.berke@gallaudet.edu

Abstract

A new extension to ELAN offers expanded n-gram analysis tools including improved search capabilities and an extensive library of statistical measures of association for n-grams. This paper presents an overview of the new tools and a case study in American Sign Language synthesis that exploits these capabilities for computing more natural timing in generated sentences. The new extension provides a time-saving convenience for language researchers using ELAN.

Keywords: ELAN, n-gram, synthesis

1. Background

In 2004 researchers at The Language Archive introduced ELAN ("Eudico Language Annotator") (Max Planck Institute for Psycholinguistics, 2013), an annotation tool that features synchronized video, audio and annotations. Its major applications include gesture studies, documentation of endangered languages and analysis of sign languages (Brugman & Russel, 2004). In subsequent years, researchers added support for synchronized motion capture (Crasborn, Sloetjes, Auer, & Wittenburg, 2006) broadened the scope of video support including enhanced time accuracy (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006), and explored automated techniques to reduce the time required to annotate media (Dreuw & Ney, 2008) (Auer, et al., 2010). In 2008, new extensions allowed users to create references from annotation systems defined in the central ISO Data Category Repository (DCR) (Sloetjes & Wittenburg, 2008). The goal is to continue to foster greater data sharing among language researchers.

Part of ELAN's appeal derives from its powerful and diverse search tools (Stehouwer & Auer, 2011). These provide an immense gamut of search granularity, ranging from finding individual annotations in local files, to accessing web-based corpora. Users can also search for n-grams within a single tier of annotation codes or for phenomena that co-occur on multiple tiers (Crasborn, Hulsbocsh, Lampen, & Sloetjes, 2013). The two main formats for search results are the concordance and frequency views. In either view, users can elect to "Show hit in transcription", which cues the linked media and annotation to the position where the hit occurs in the ELAN annotation file.

Statistical services available include frequency counts for search queries. The "Statistics for multiple files" search also includes basic descriptive statistics for hits within a tier, including the duration minimum, maximum, mean and median. For further analysis, ELAN provides export capability so that researchers can export the raw search results for further study.

The rest of this paper is organized as follows: Section 2 describes the potential of using ELAN as a resource to develop sign synthesis systems; section 3 is an in-depth analysis of a sparse-key synthesis system, and the benefits of n-gram analysis for such a system; section 4 presents a new software module for ELAN that implements several n-gram statistical tools; section 5 discusses the application of the results of an n-gram analysis to sign synthesis, and section concludes with results and future work.

2. Using ELAN in transcribing American Sign Language (ASL) for animation

An additional application for ELAN is the automatic generation of sign language (Efthimiou & Fotinea, 2007), (Efthimiou E. , et al., 2009). For artists animating individual signs, having ELAN and a corpus conforming to the ELAN format is an invaluable resource. To take advantage of the user interface and search capabilities of ELAN, we wrote a tool based on Christian Vogler's software library (Vogler, 2013) to convert the ASLLRP corpus (Neidle & Vogler, 2012) from its native SignStream format, as there are currently no corpora of ASL publically available for ELAN.

Corpora in ELAN format offer three distinct advantages for artists. First, well-planned corpora are application neutral, in contrast to exemplars recorded for a specific purpose. Such special purpose exemplars can prove problematic. We have found that, if signers are aware of a study's purpose, they will tend to over-articulate to demonstrate all of the subtleties related to the phenomenon under study. Unfortunately, these overarticulations often find their way into the animations. Second, signs in a corpus are performed in context and there are typically multiple examples of individual signs. Artists gain greater insight when provided with multiple examples in context. Third, ELAN provides extremely precise and quick access to recorded signs in a corpus through the “Show hit in transcription” option. As a result, artist productivity is increased by obviating the need for recording exemplars or for Web searches.

3. Sign synthesis requirements

Our sign generation system produces sentences from a library of pre-animated signs, which consist of sparse key frame data. The system’s capabilities include conjugating agreement verbs, prosodic punctuation, and co-occurring nonmanual signals, both facial and spinal. Figure 1 is a close-up of the system’s avatar demonstrating a co-occurring nonmanual signal and pragmatic, and Figure 2 is a still frame demonstrating pointing to a previously-established referent. This approach has the advantage of extremely fast real-time synthesis speeds, yielding rendering rates that are typically three times conventional video frame rates.

Figure 1: Close-up of signing avatar asking a question with pleasant affect.

At present, the library has sufficient data and procedures to generate sentences that are well-formed grammatically, but does not have the necessary data to create movement that is convincingly lifelike. A major contributing challenge is the difficulty of computing the intervals between signs. In typical discourse, the length of individual signs fall in the range of 0.5 to 1.0 seconds, but little is known about characterizing the temporal spacing between signs.

Two research efforts have considered the coarticulatory effects on sign location in ASL (Grosvald, 2009) (Mauk, 2003), which studied whether a sign having a high position, would be influenced by preceding and succeeding signs having a low position. Both investigations found that the position of non-contact signs (signs where the hand(s) are not touching the body) are influenced by the position of the preceding and succeeding signs. Researchers exploring Langue des Signes Française (French Sign Language) found similar behavior (Segouat, 2010).

All three of these studies looked at the change in location, but did not consider the change in temporal spacing. The question is, “Does the inter-sign temporal spacing change due to location?” Consider the two phrases “FATHER WANT” and “CHILDREN WANT”. FATHER is located on the forehead, whereas WANT and CHILDREN are both signed in neutral space in front of the body at midriff height. What we want to determine is if there is a larger inter-sign duration for the phrase FATHER WANT, because FATHER and WANT have locations that are farther apart.

French researchers (Awad, Courty, Duarte, Le Naour, & Gibet, 2010), (Braffort, et al., 2010) have explored this topic in regards to generating Langue des Signes Française (LSF). In ASL, however, researchers have focused solely on temporal spacing to delineate clauses and sentences (Huenuerfauth, 2008) (Adamo-Villani, Hayward, Lestina, & Wilbur, 2010). Their timing was based on heuristics and did not address the intervals between lexical items interior to a clause.

Figure 2: Referring to a previously-established protagonist.

To improve automatic ASL generation by incorporating data-driven inter-sign timing requires data derived from bigram analyses to determine the natural intervals between signs. However, simply storing the data associated with every possible bigram as an entry in a lookup table is not currently possible for real-time generation because such a table can grow so large as to become unwieldy. Instead, a practical tradeoff is to store only a subset of the possible bigrams. However, simply choosing the most frequently occurring bigrams in a corpus may not accurately represent the most important ones, especially when using small corpora (Dunning, 1993). Dunning classifies a corpus of 33,000 English words as "small". The ASLLRP has approximately 9,000 annotations in the main gloss tier, which tends to put it in the small category as well. For this reason, more analysis is required than the current version of ELAN can offer, as it only offers frequency counts.

4. A new extension to ELAN: an n-gram analysis tool

The new extension presented here offers the time-saving convenience of built-in n-gram statistical tools, similar to ANVIL (Kipp, 2010), in the environment of ELAN's interface and data sharing capabilities. Previously, performing an n-gram analysis on a corpus required exporting data from ELAN. Reformatting the exported data and/or developing the methods for computing n-gram analyses can be time consuming. To address this, we built an n-gram analysis extension to ELAN, accessible through the "Multiple File Processing" entry in the File menu. What follows is a description of the new capability in ELAN (Berke, 2013).

The N-gram analysis is located in the "Multiple File Processing" submenu of the File menu (Figure 3).

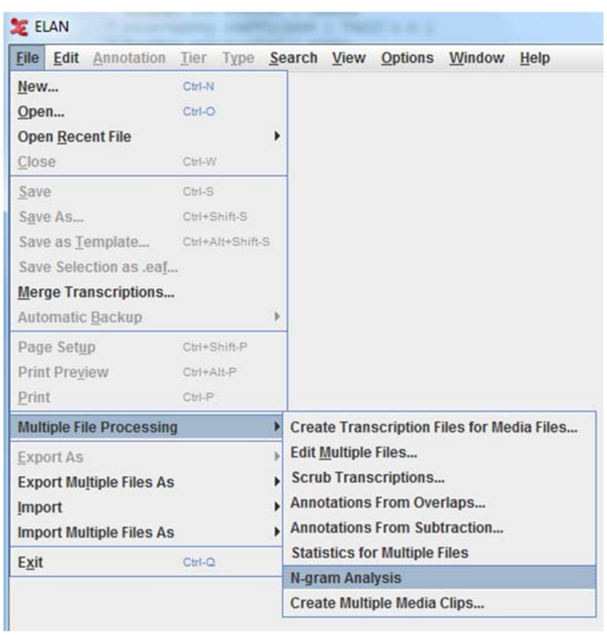

Figure 3: Location of the N-gram analysis in ELAN.

After selecting the n-gram analysis menu item, a user will see a new dialog window pop up that contains the various options for the search (Figure 4).

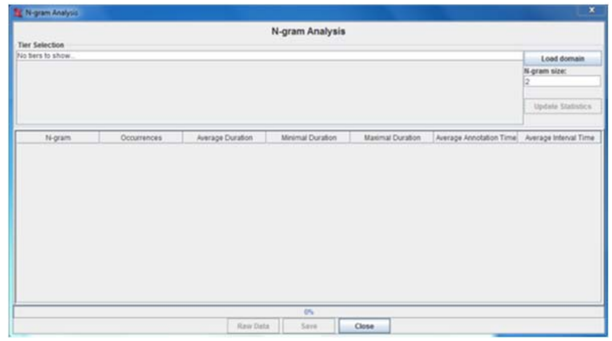

Figure 4: Main N-gram analysis window.

The first step is to select the search domain via the standard ELAN "Load domain" window where the user can specify a list of files or directories. Once the domain is designated, the dialog will display a list of tiers contained in the domain (Figure 5). This functionality assumes that all annotation files in the domain have an identical set of tiers. The software then loads the first file in the domain to extract the tiers and then displays the results in the window.

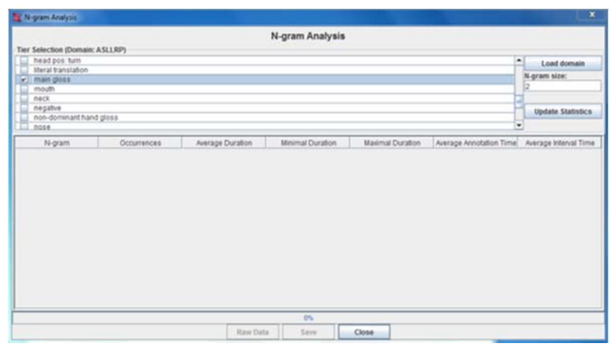

Figure 5: N-gram analysis window displaying possible tiers for searching.

The user then enters the N-gram size in the textbox. The software can handle a N-gram of any size, however the contingency table analysis can only be performed on bigrams. With an N-gram size entered, the user can click the "Update Statistics" button to start the search. The annotations are extracted from the files, the N-grams created from them, and the results are finally collated by N-gram for statistical analysis.

When the search is done, a report window will pop up displaying some search-related statistics. This can be used to check the integrity of the search. A sample report is seen in Figure 6.

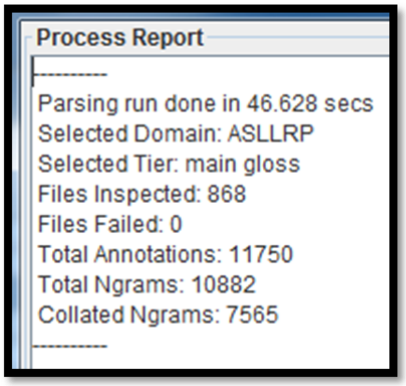

Figure 6: N-gram analysis report window.

Figure 7 is a screen shot of the user interface. To make the new extension easy to use, we designed the interface to be as consistent as possible with other ELAN search dialogs.

Similar to the existing n-gram search, users can designate a search domain. However, the new extension augments the n-gram search results, by adding the n-gram measures of association as described in (Banerjee & Pedersen, 2003). These are listed in Table 1. Users can choose the metrics most appropriate to their work. In addition, the extension still offers users the option to export the results as a tab-delimited file for further study.

Chi-squared	Dice coefficient	Jaccard coefficient
Fisher exact two tailed	Fisher exact left sided	Fisher exact right sided
Phi coefficient	Pointwise mutual information	Log-likelihood
True mutual information	Odds ratio	Poisson-Stirling measure
	T-score	

Table 1: N-gram metrics of association

5. Application in ASL synthesis

The goal of this work was to create a data-driven approach to timing aspects of coarticulation at the lexical level, similar to that of (Awad, Courty, Duarte, Le Naour, & Gibet, 2010) and (Braffort, et al., 2010), but for American Sign Language and utilizing the "sparse key" approach. For this we used the new n-gram analysis extension to extract the 1100 most important bigrams based on the log likelihood statistic described by (Dunning, 1993). Depending on where the bigram occurred, we used the maximal duration for end-of-sentence punctuation, the third quartile duration for clausal punctuation, and the median duration otherwise.

6. Results and Future Work

Incorporating data-driven timing has had no noticeable effect on rendering speeds in the real-time system due to the limited number of bigram records. We will demonstrate the improvement over the previous system with side-by-side animations of selected ASL sentences. The n-gram extension to ELAN is available as a patch to version 4.6.1 at http://asl.cs.depaul.edu/software/ELAN-patches-Berke.zip. Documentation is available at http://asl.cs.depaul.edu/software/ngramELANtechReport.pdf .

Ultimately, we would like to supplement the data-driven timing with a heuristic, to accommodate those cases where the bigram is not in our table. This will require analysis of sign location, which are not coded in the corpus we are currently using. We will need to identify additional corpus resources to complete this task.

Future work includes refining this approach through analysis of the English transcription tier as well as applying this approach to the timing of nonmanual signals.

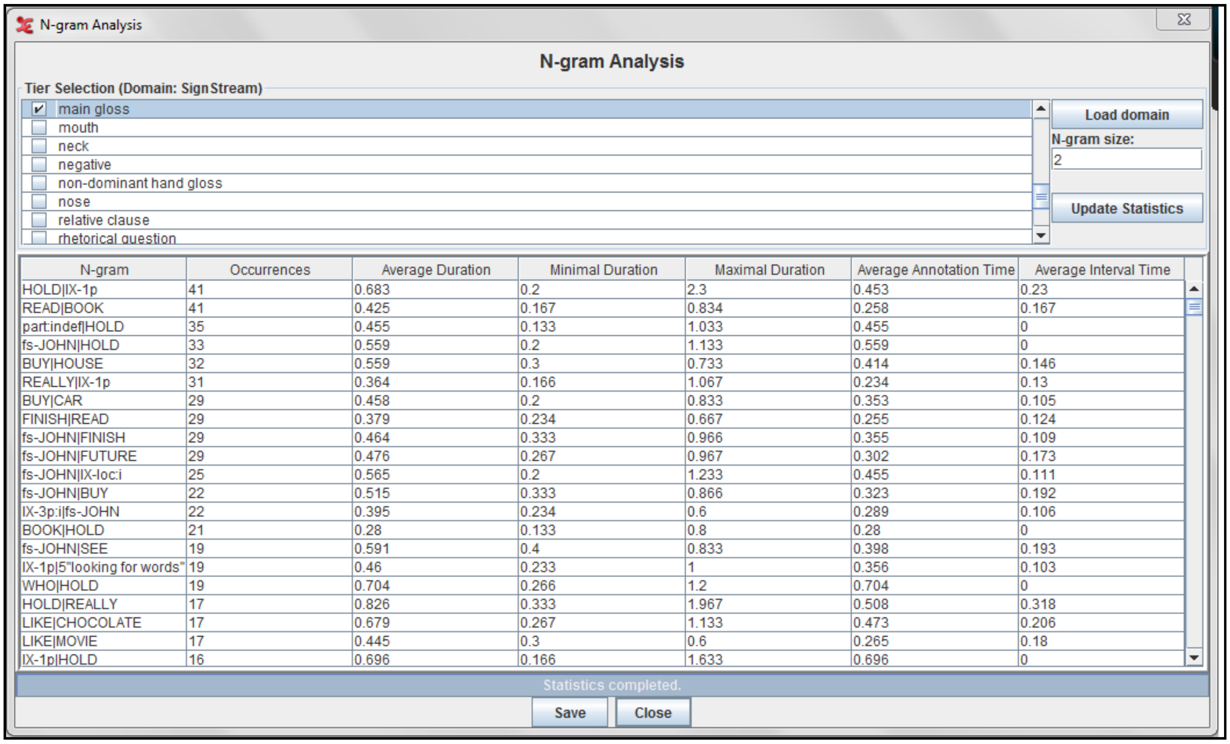

Figure 7: Search results from new N-gram analysis. Not all columns are visible.

7. Acknowledgments

This work would not have been possible without the generous support of CRA-W's Distributed Research Experiences for Undergraduates (DREU) and the Alliance for Access to Computing Careers (AccessComputing), sponsored by the NSF.

8. References

Adamo-Villani, N., Hayward, K., Lestina, J., & Wilbur, R. (2010). Effective animation of sign language with prosodic elements for annotation of digital educational content. *ACM SIGGRAPH 2010 Talks*, (p. 39).

Auer, E., Russel, A., Sloetjes, H., Wittenburg, P., Schreer, O., & Masnieri, S. (2010). ELAN as Flexible Annotation Framework for Sound and Image Processing Detectors. *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010)* (pp. 890-893). Paris: ELRA.

Awad, C., Courty, N., Duarte, K., Le Naour, T., & Gibet, S. (2010). A combined semantic and motion capture database for real-time sign language synthesis. *Intelligent Virtual Agents*,, 432-438.

Banerjee, S., & Pedersen, T. (2003). The design, implementation, and use of the Ngram statistics package. *Computational Linguistics and Intelligent Text Processing*, 370-381.

Berke, L. (2013). *Empowering ELAN with N-gram Analytics for Corpora*. DePaul University, School of Computing, Chicago. Retrieved from http://asl.cs.depaul.edu/software/nGramELANtechReport.pdf

Braffort, A., Bolot, L., Chételat-Pelé, E., Choisier, A., Delorme, M., Filhol, M., . . . Devos, N. (2010). Sign Language corpora for analysis, processing and evaluation. *Proceedings of the 7th Int. Conference on Language Resources and Evaluation (LREC 2010).*

Brugman, H., & Russel, A. (2004). Annotating Multi-media/Multi-modal Resources with ELAN. *Proceedings of the 2nd Workshop on the Representation and Processing of Sign languages: Lexicographic matters and didactic scenarios* (pp. 82-87). Paris: ELRA.

Crasborn, O., Hulsbocsh, M., Lampen, L., & Sloetjes, H. (2013). New multilayer concordance functions in ELAN and TROVA. *Proceedings of the Tilburg Gesture Research Meeting [TiGeR 2013].*

Crasborn, O., Sloetjes, H., Auer, E., & Wittenburg, P. (2006). Combining video and numeric data in the analysis of sign languages within the ELAN annotation software. *Proc. LREC 2006 workshop on representation & processing of sign language* (pp. 82-87). Paris: ELRA.

Dreuw, P., & Ney, H. (2008). Towards automatic sign language annotation for the ELAN tool. *LREC Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*. Paris: ELRA.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, *19*(1), 61-74.

Efthimiou, E., & Fotinea, S.-E. (2007). GSLC: creation and annotation of a Greek sign language corpus for HCI. *Universal Acess in Human Computer Interaction. Coping with Diversity*, 657-666.

Efthimiou, E., Fotinea, S.-E., Vogler, C., Hanke, T., Glauert, J., Bowden, R., . . . Segouat, J. (2009). Sign language recognition, generation and modeling: A research effort with applications in deaf communication. In C. Stephanidis, Ed., *Universal Access in HCI* (pp. 21-30). Berlin/Heidelberg: Springer.

Grosvald, M. (2009). *Long-Distance Coarticulation: A Production and Perception Study of English and American Sign Language.* Doctoral Dissertation, University of California, Davis, Davis, CA. Retrieved February 21, 2013, from http://linguistics.ucdavis.edu/pics-and-pdfs/MGrosvaldDiss.pdf

Huenuerfauth, M. (2008). Evaluation of a psycholinguistically motivated timing model for animations of american sign language. *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility* (pp. 129-136). ACM.

Kipp, M. (2010). Multimedia annotation, querying and analysis in ANVIL. *Multimedia information extraction, 19*.

Mauk, C. (2003). *Undershoot in two modalities: Evidence from fast speech and fast signing.* Doctoral Dissertation, The University of Texas at Austin, Austin, Texas. Retrieved February 14, 2013, from http://repositories2.lib.utexas.edu/bitstream/handle/2152/762/maukce036.pdf?sequence=2

Max Planck Institute for Psycholinguistics. (2013, September 19). *ELAN.* Retrieved from The Language Archive: http://tla.mpi.nl/tools/tla-tools/elan/

Neidle, C., & Vogler, C. (2012). A new Web interface to facilitate access to corpora: Development of the ASLRP Data Access Interface. *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon.* Istanbul, Turkey.

Segouat, J. (2010). *Modélisation de la coarticulation en Langue des Signes Française pour la diffusion automatique d'informations en gare ferroviaire à l'aide d'un signeur virtuel.* Doctoral Dissertation, Université Paris Sud, Orsay, France. Retrieved February 21, 2013, from http://hal.upmc.fr/docs/00/60/21/17/PDF/these-segouat2010.pdf

Sloetjes, H., & Wittenburg, P. (2008). Annotation by category - Elan and ISO DCR. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008).* Paris: ELRA.

Stehouwer, H., & Auer, E. (2011). Unlocking Language Archives Using Search. *Proceedings of the Workshop on Language Technologies for Digital Humanities and Cultural Heritage* , (pp. 19-26).

Vogler, C. (2013, August 3). *SignStream-XML Parser.* Retrieved from ASLLRP: http://www.bu.edu/asllrp/ncslgr-for-download/sinstream-parser.zip

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). Elan: a professional framework for multimodality research. *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)* (pp. 1556-1559). Paris: ELRA.