Beyond Translation -- new possibilities for reading in a digital age (DH Level 3)

Our goal in this project is to promote a fundamental change in how human beings view translations and the cultures of which their original source text is a product. In a period where marketers and the popular press focus on artificial intelligence, machine translation has become increasingly important. But for us as humanists in general (and as philologists, focused on understanding the human record as deeply and broadly as possible), translation is only one instrument among many and, if seen as a final goal, a barrier to understanding others whose parents spoke to them in languages other than our own and who look back upon cultural traditions very different from those of most Americans. At some level translation obscures more than it reveals, offering a deceptively simple equivalent. Even when we translate between European Spanish and North American English -- languages that evolved together within Europe -- research shows that terms such as "shame" and "vergüenza" mean fundamentally different things to speakers in the United States and Spain. Our goal is to allow readers to push past the translation and to explore for themselves how original sources frame human experience.

Enhancing the Humanities: Several ongoing changes motivate the work proposed here. **First**, researchers on reading in a digital age have reported alarming consequences from much reading on digital devices. Maryanne Wolf, a leading expert in the effects upon reading, summarized many of these concerns in an August 2018 article published in the *Guardian*: "Skim reading is the new normal. The effect on society is profound".[1] She reports that her own research "depicts how the present reading brain enables the development of some of our most important intellectual and affective processes: internalized knowledge, analogical reasoning, and inference; perspective-taking and empathy; critical analysis and the generation of insight. Research surfacing in many parts of the world now cautions that each of these essential 'deep reading' processes may be under threat as we move into digital-based modes of reading." The work proposed here exploits digital media to support something profoundly different from skimming by challenging audiences to pause and explore sources in languages that they may never have studied. Our goal is to enable and to foster a form of deep and active reading that has its roots in traditional close reading but that, through the use of digital media, allows the emergence of new, deep and open-ended forms of intensive reading. We seek to foster a new and far more dynamic relationship between translations and the source texts that they represent.

Second, there is the contrast between the increased prominence of English, on the one hand, as the language of business, scholarship, and popular entertainment in many parts of the world and, on the other, the radical increase in access to authentic materials in foreign languages. A generation ago, language learners went to language labs to listen to native speakers on tape. Now, streaming media such as YouTube and social media of all kinds make it possible to

---

[1] https://www.theguardian.com/commentisfree/2018/aug/25/skim-reading-new-normal-maryanne-wolf

experience dozens, if not hundreds, of languages. Netflix, for example, now buys global rights to, or produces itself, original content produced in languages such as Hindi, in Turkish, in various forms of Spanish and other languages but the accompanying subtitles are essentially unchanged from those available for decades in film.

Our goal is to make the source in an unknown language a challenge and an opportunity. Where films have, for generations, provided subtitles, media such as YouTube now often include both a transcription of the original text and an accompanying English translation.[2] We do so by exploiting digital media to produce exhaustive links between source text and translation at the word and phrase level, as well as a growing body of linguistic annotation to open up the original source text so that audiences can immediately begin to pick it apart and can even begin the process of learning the language. We can now align source texts and translations at the word and phrase level where the two closely correspond while, equally important, revealing the gaps where a translation has left out, augmented, or so changed the original sense that alignment is not possible. The requested funding would support a range of complementary outcomes:

- Reviews that engage readers in the original source texts at a level that has never before been feasible.
- Automatically generated alignments between source text and translation and linguistically sophisticated reading tools that allow readers to explore the relationships between translations and source texts in languages that they have not (yet) studied.
- A Dynamic Lexicon that exploits aligned translations and automated linguistic analysis to provide extensive new lexicographic information and semantic exploration of (in the three years of work proposed here) Greek and Latin, with exploratory work on German, Arabic, and Persian.
- A new generation of born-digital aligned translations that are designed not only to support advanced study but that also transform the way in which learners produce, receive feedback on, and learn vocabulary from texts in languages that they are learning.
- Integration of translation alignment within pedagogy of historical languages, with students producing aligned translations from the first weeks of study, both to practice their own linguistic knowledge and to generate new quizzes on vocabulary.

Extensive prior activity, including print and digital scholarship and an enduring international audience make classic works such as the Homeric Epics and the Persian poetry of Hafez practical starting points. The applications of the work proposed here are broad -- there is the potential for fundamental change in how the public sees and works with translations and the languages from which they are drawn -- and much activity may take place in social networks and big tech, A great deal of effort has already gone into creating corpora with exhaustive linguistic annotation, with roughly one million analyzed words available under an open license

---

[2] E.g., under the "show more" link, Middle-high German and English at https://www.youtube.com/watch?v=7I0yQ9T8nNI; Latvian and English at https://www.youtube.com/watch?v=v5hDhkVEboY.

for Greek and Latin in particular[3]. NEH support would allow us to focus on the problem of aligning English translations with the source texts and these linguistically analyzed corpora. In so doing, we build upon the tangible outcomes from more than a decade of work, described below and funded by the NEH, as well as the National Science Foundation, the Fund for the Improvement of Postsecondary Education, the Institute for Museum and Library Services, and the Mellon Foundation in the US, and by the Alexander von Humboldt Foundation, the DFG, and the Ministry for Research, and Education in Germany. All code, data and publications from the proposed work will be available for third-party work. At the same time, as part of a new version of the Perseus Digital Library (PDL), these services will reach an audience of more than a quarter million users per month.

Environmental Scan

The work proposed here both utilizes technologies with which Perseus has an extensive development history (e.g. morphological analysis, treebanks, advanced reading support) and also explores a number of newer digital methodologies (e.g. user annotations and alignments, automatic translation alignment).   Several recent scholarly overviews of the field digital classics have provided a look at many of the technologies we plan to implement (see Buchanan 2015; Coffee & Bernstein 2016). As this project is particularly concerned with the tools of digital philology and digital reading for classical languages and how they might be used by wider audiences, we will also consider recent research by Dilley (2016) and Barker and Terras (2016). Recent digital research in Ancient Greek morphology (Smith 2016) as well as in digital tools and services for automatic morphological analysis for other languages such as Coptic (Schroeder & Zeldes 2016) will also be reviewed.

While the Beyond Translation project will focus specifically on user contributions and annotations in the form of translation alignment and treebank data, other projects that have made use of collaborative annotation models for other types of data will still inform our research. In particular, the ongoing work of the Pelagios Commons,[4] which provides "*community* & *infrastructure* for *Linked Open Geodata* in the *Humanities*" with a particular focus on the ancient world, and its well-supported annotation tool Recogito 2 (Simon et al. 2017) provides a model we will consider as our project progresses. Similarly, a variety of recent scholarship has examined new ways to engage a broader audience within digital classics and how to solicit a variety of different types of user contributions. The 2016 monograph *Digital Classics Outside the Echo Chamber: Teaching, Knowledge Exchange & Public Engagement* (Bodard and Romanello 2016), includes a number of chapters covering projects such as Perseids (Almas and Beaulieu 2016), Ancient Lives (Brusuelas 2016), and the Eagle Project (Orlandi 2016) that all offer different methods and models to build and sustain user communities and new digital modes of learning (Rydberg Cox 2016).

---

[3] E.g., https://perseusdl.github.io/treebank_data/; https://proiel.github.io/; https://itreebank.marginalia.it/; https://perseids-project.github.io/harrington_trees/.
[4] http://commons.pelagios.org/

Although the PDL has long presented translations in alignment with source texts, the practice of using parallel texts or in some cases parallel corpora to support language learning is fairly limited within educational or historical digital libraries. One of the most common uses of parallel texts or corpora is to support tasks such as machine translation in natural language processing (NLP) research.[5] While a large number of digital libraries include translations of primary source texts (e.g. Cervantes Virtual Library,[6] Europeana[7]), they are typically presented in isolation and not in alignment with that source text. One notable and fairly new exception is the Archaeology of Reading[8] project, which has included aligned translations for several of its classical texts. In addition, when translation alignment is used within digital humanities or digital library projects, the goal is most typically to design tools (Bozzi 2013) that assist scholars in complicated tasks such as version comparison (Cheesman et al. 2017, Cheeseman and Roos 2017) or advanced translation studies (Abzianidze et al. 2017) rather than in assisting students just beginning with a new language. For example, the Vicoglossia system has been designed (Maiaetsky et al. 2018) to help philological scholars align early 20th century Russian philological texts with English or other language translations. Other relevant work has documented the creation of digital tools that allow scholars to compare multiple or in some cases hundreds of translations of a given work, including Homer's *Odyssey* (Bizzoni et al. 2017), Mark Twain's *Huckleberry Finn* (Fraisse et al. 2018), Kafka's *Metamorphosis* (Cap, Rösiger, and Kuhn 2015), and the Bible (Christodouloupoulos and Steedman 2015).

History of the Project

Background on Perseus: According to Google Analytics, the PDL website served 3.1 million users in the twelve months ending December 19, 2018. Tufts University has provided a home for the Perseus Project since 1992 and has supported the PDL website since 1995. The Winnick Family Chair for Technology and Entrepreneurship supports PI Crane, while Tufts supports two positions dedicated to Perseus: Project Manager Lisa Cerrato and Digital Librarian Alison Babeu. In addition, Tufts supports the ongoing costs of maintaining Perseus. While Tufts cannot finance substantial new development, it is well-positioned and indeed eager to maintain the work proposed herein. The Perseus website was designed in 2004 and a new Perseus has been under development. NEH support would help us build on work that we already have done with aligned translations. This work builds, in turn, upon backend support for the Canonical Text Services (CTS) data model[9] and a new, extensible front-end, named the Scaife Digital Library Viewer.[10]

Support for the CTS data model: On the back end, Perseus has adopted the Canonical Text Services data model, a framework that allows us to cite specific spans of text within particular editions of a work. Perseus and the Open Greek and Latin Project have made more than 40

---

[5] For a detailed analysis of this use, please see Véronis (2000), for NLP uses of translation corpora see Kirk et al. (2018).
[6] http://www.cervantesvirtual.com/
[7] https://www.europeana.eu/portal/en
[8] https://archaeologyofreading.org/
[9] http://folio.furman.edu/projects/citedocs/cts/
[10] https://scaife.perseus.org/

million words of Greek and Latin available on GitHub[11] in CTS-compliant TEI XML. At the same time, collaboration between the Alexander von Humboldt Chair of Digital Humanities at Leipzig and the Mellon-funded Perseids Project at Tufts produced the CapiTainS[12] software suite and guidelines for citable text, an open-source, scalable backend that helps produce and serve texts via the CTS protocol. Development of the Scaife Viewer, a new front-end for Perseus: Funding from Harvard's Center for Hellenic Studies (CHS) and the Alexander von Humboldt Foundation allowed us in March 2018 to produce an initial release of the Scaife Digital Library Viewer, a new reading environment for Perseus. This viewer provides a front-end to the CTS texts served by the CapiTainS server. It includes a completely redesigned backend and more modular front-end, with support for reading and browsing sources in a variety of languages. To emphasize our goal to provide generalized solutions and to commemorate a pioneer in Digital Classics, we have called this front-end the Scaife Digital Library Viewer, in memory of the late Ross Scaife, 1960–2008. Support for Beyond Translation will allow us to transform the role that translations play within the new Perseus that is built on the Scaife Viewer and that draws on an increasingly rich variety of annotations (such as the exhaustive morpho-syntactic annotations from the Greek and Latin Treebanks).

Prior work relevant to working with translations: The intellectual impetus for PI Crane's work on this project began in fall 1975. As a first-year college student, Crane learned from classmates that Professor Gregory Nagy had students with no knowledge of Greek engage directly with the Greek text of the Homeric epics and pose their own questions about what that text meant. There were no interactive digital tools. Students spent roughly one class learning the Greek alphabet and they then used the 19th century print concordance and English translations of the *Iliad* and *Odyssey* where the line numbers lined up precisely with those in the Greek text. Given a general sense of what a Greek word meant (e.g., *aidôs*, "shame"), if they knew that that word showed up in a line of Greek, they could usually determine the corresponding English translation and, in so doing, they, like professional philologists, built up their own inductive models of what words meant and how concepts were shaped in a language and culture that differs fundamentally from that of their own.

A decade later, this bilingual semantic search and exploration motivated much of the earliest work on Perseus. We developed a morphological analyzer that would enable linguistically smart searches so that readers could enter a single dictionary entry and locate all possible inflected forms of that word in a corpus. When combined with a facing page translation, this smart search allowed users with no Greek to implement the Nagy class exercise and begin exploring concepts in Greek on their own. Crane tested this in a class on fifth century BCE Greek literature in English translation, with (in his view) strong results: the students couldn't, of course, always align the English to the Greek, but they were successful enough in their alignments to develop a range of interesting ideas despite not knowing the Greek. In the late 1980s, Professor Gary Marchionini, then of the University of Maryland, had external reviewers assess the Greek

---

[11] https://github.com/PerseusDL; https://github.com/opengreekandlatin/; https://github.com/OpenGreekAndLatin/First1KGreek.
[12] http://capitains.org/

abilities of students who had written papers using the Nagy method with this early, CD-Rom based version of Perseus (Marchionini and Crane 1994). The reviewers assumed that the students had advanced expertise in Greek. These judgments provided external evidence that readers with no knowledge of a language could exploit coarsely aligned translations to explore the semantics of a language such as Greek. Starting in 2013, Crane, working with Anna Krohn, returned to this method. They added English translation exercises into a course on Greek literature where students compared two English translations to the Greek, explored the meanings of Greek words, and exhaustively studied short passages in the Greek original (Krohn and Crane 2014).

Exhaustive linguistic annotation, describing the morphological and syntactic analyses for every word in a text, provides another component to the work proposed here. These annotated databases (called "treebanks" because the syntactic data is stored in a hierarchical graph that looks like a tree when displayed) were designed to support linguistic research (e.g., locate all instances where a particular verb takes objects in the dative) but their level of detail also helps readers decode the function of each word in a sentence. Perseus developed treebanks not only to support research on Greek and Latin but also to support readers. Treebanks offer readers comparing translation and source text a tool by which to identify much more precisely the function of each word in the source text. NSF provided planning support for a Latin Treebank in 2006[13] and the Alpheios Project[14] provided support in 2008 that allowed Perseus to develop a Treebank for Ancient Greek. As noted above, roughly one million words of Latin are available in openly licensed treebanks. The translingual reading environment proposed here will build upon treebank data.

David Bamman[15] developed the Latin treebank. David used automated methods for parallel text alignment to create word and phrase level links between our Greek and Latin source text and translations. He combined these automatic alignments with automatically generated syntactic analyses to detect (1) different word senses (by seeing different translation equivalents of Greek and Latin words in English, e.g., *oratio* as "speech" vs. "prayer") and (2) basic syntactic data (e.g., the most common subjects and objects of verbs, the most common noun/adjective combinations). Bamman received NEH funding for the Dynamic Lexicon project[16] and his work was based on access to English translations for 5 million words of Greek and 3 million words of Latin (Bamman and Crane 2008). Between 2014 and 2016, the European Social Fund and the Alexander von Humboldt Foundation provided support to expand available Greek and Latin to more than 40 million words with 50 million words of accompanying English translation.

Where Bamman focused on automatic alignments, in 2008, Alpheios began work on a front end by which readers could manually align words and phrases in source text and translation. In 2015, Tariq Yousef and Maryam Foradi, PhD students at Leipzig, developed a reading

---

[13] BCS-0616521: https://www.nsf.gov/awardsearch/showAward?AWD_ID=0616521
[14] https://alpheios.net/; the actual funding came from the Cantus Foundation.
[15] Then at Perseus, now an assistant professor at the Berkeley I-School.
[16] PR-50013-08: https://securegrants.neh.gov/publicquery/products.aspx?gn=PR-50013-08

environment[17] with the poems of Hafez, a twelfth century Persian poet. Maryam aligned this 70,000 word corpus at the word/phrase level with an English translation. This system also provided an overview of how words were translated in different contexts as well as a bilingual search function so that readers could examine all the contexts where a word appeared in one language and what it corresponded to in the other (e.g., readers could search for English words and see the Persian equivalents and vice versa). In 2016, Tariq developed a newer alignment editor[18] with a more advanced front end than was available from the original Alpheios alignment editor. As of December 2018, this system contains alignments involving not only Greek, Latin and English but 31 other languages as well produced by over 220 users. Tariq's annotation environment provides us with a model for services and future Perseus enhancements.

Alignments have also been integrated into ongoing Greek language classes. Students in Crane's classes are required to produce translations that are aligned to the source texts. In so doing, students can more precisely track how their English relates to the source texts. At the same time, students produce born-digital aligned translations that can provide more precise reading support for others, training data for improved automatic alignment of larger corpora, and raw materials for language learning (such as flashcards).

In addition to the above referenced work, Maryam Foradi's in-progress PhD poses two questions about translation alignment. **First**, Maryam compared how accurately two different groups generated alignments between the Persian and a German translation. She determined that both groups produced comparably accurate alignment data and thus provided concrete evidence to demonstrate how well readers could connect a translation to words in an unfamiliar source text. **Second**, Maryam asked if the students learned any Persian from the incidental exposure as they aligned Persian to German. She discovered that annotators actually learned Persian vocabulary in the course of alignment. While flashcards are more effective for short term learning, two months later students remembered Persian words that they had aligned just as well as those who had used flashcards. One of our goals in the work proposed here is to build on Maryam's research, integrating the effects on readers of using treebank as well as alignment data.

Finally, in fall 2018, Sophia Sklaviadis joined the PhD program in Computer Science at Tufts to apply computational methods to Ancient Greek. She led a student team that applied classical statistical machine translation (SMT) methods to several translations of the *Odyssey*. The group collected manual alignments at the sentence level from the original Ancient Greek to English, French, Italian, Spanish, Modern Greek, Persian, and Chinese. From these sentence alignments the group used the open source SMT software GIZA++ to automatically generate word-level alignments. They tested various standard automatic alignment models (IBM 1, IBM3, and HMM:Och 2003) across language pairs and confirmed the superior performance of the HMM model (~67% accuracy). One particularly interesting outcome of this work is that only the

---

[17] http://www.divan-hafez.com/
[18] http://www.ugarit.ialigner.com/

HMM model's output improved significantly with additional data. Sophia is planning to generate data from integrated neural machine translation models to include in this study.

Work Plan

NEH support would allow us to integrate, and to scale up, the results described above into an environment which will, we believe, be greater than the sum of its parts. The work plan focuses on three resources: (1) a larger bilingual corpus where parallel text alignment algorithms have automatically aligned source texts and translations and where morpho-syntactic parsers have automatically suggested the form and syntactic function of each word in the source texts (so that readers can also work directly with the source text); (2) a smaller corpus of texts where human editors have created alignments and produced morpho-syntactic analyses by hand; (3) a reading environment that provides lightweight access to the morpho-syntactic data and to the alignments. The automatically produced and the human curated annotations feed into each other. Readers may correct automatically generated annotations. The automated algorithms can then use the corrected annotations as training data to improve subsequent automated analysis. The work plan is designed to support a use case such as the following:

A reader explores the meaning of a particular word. The reader may begin with a word in a particular passage in an English translation (e.g., "anger") and ask to see the corresponding term in the source text (e.g., *mênis* in the first line of the *Iliad*). The reader then asks whether that translation word corresponds to different words in other passages of the source text (e.g., in the *Iliad*, the Greek word *cholos* can also be translated as "anger"). The reader then examines a number of passages where the different words appear and inductively develops a model for how the source text words differ from each other (in the *Iliad*, *cholos* is the standard word for anger while *mênis* is the word used for the anger of gods and of Achilles).

The reader uses a range of digital reading aids to work directly with the source text. Core aids include (1) data about the form of inflected words (e.g., Latin *fecit* is a 3rd singular perfect indicative active of the verb *facio*, along with a general definition such as "to do, to make"); (2) links to one or more online lexica (e.g., dictionary entries with extended information about the meaning of *facio*); (3) a list of words used to translate that particular form (e.g., *fecit*) as well as any form of that dictionary entry (e.g., all forms of *facio*) and the ability to browse the passages where that word appears (e.g., see every instance of *fecit* or of any form of *facio* and see how it was translated).

At this point, readers should have the ability to create a source book in which they can organize passages into different categories reflecting word senses or any other category relevant to their work. In so doing, readers need the ability to correct and augment existing annotations (including offering alternative analyses to those provided by human annotators). This process inevitably requires the ability to modify the translation and even the source text (scholars regularly identify as quoted modern language translation as "based on" a pre-existing translation, indicating that they have made changes). In this annotation environment, we retain

the original translation and offer subsequent readers the ability to see where and how the original translation was modified.

Finally, the reader has the opportunity to offer corrections to existing annotations for publication in Perseus and to publish dynamic source books of various forms (e.g., a monograph-length study of a particular word, a sourcebook on medical terminology or gender terms or some other topic). The source books contain links back to the original sources so that readers can explore the contexts from which they are drawn.

Our goal is to empower as wide a community as possible of these reader/editor/authors. On the one hand, our goal is to support specialist researchers in a field (e.g., an expert in Greek drama who carefully considers how different editions reconstruct particular passages) as well as members of the general public. (The widely used Landmark Thucydides,[19] for example, which contains exhaustive annotations about people, places and topics, was produced by Robert Strassler, a businessman with no formal training in Thucydides).

Tasks: (1) The software development firm that developed the Scaife Viewer, Eldarion,[20] will integrate the following features into the Scaife Viewer:
1. Incorporation of display of text alignment and morpho-syntactic analyses into the reading environment user interface. (months 1-3)
2. Implementation of user interface for display of all words used to translate a particular source word and all source words from which a particular translated word has come. (months 4-6)
3. Implementation of feature for authenticated users to make corrections to alignment and morpho-syntactic analyses in a non-destructive way, and to manage their corrections (including the visibility of those corrections to others and contribution back to Perseus, (months 7-9).
4. Implementation of a feature for authenticated users to collect passages, source language words, and translated words into a source book (months 10-12).
5. Implementation of feature to export source book data (month 13).
6. Ongoing support of the Scaife ecosystem and community (months 1-36).
7. Consulting with Crane and Sklaviadis to make sure any software they develop can be integrated and sustained within the Scaife Viewer(months 1-36).

(2) Sklaviadis will be responsible for automatic alignment of sources and translations:
1. Application of automatic alignment with Giza++ to 50 million words of English, 20 million words of German, 20 million words of French, and 8 million words of Italian translation for Greek and Latin that have been converted into TEI XML since 2014 (focusing on the output of the Hidden Markov Model Alignment Model, Vogel 1996); (months 1-6).

---

[19] https://www.worldcat.org/oclc/39825812
[20] https://eldarion.com/; James Tauber, who runs the company is also a student of Greek and has two decades of digital work on the New Testament.

2. Evaluation and (if appropriate) application to alignment of deep learning/neural machine translation (NMT) methods that have emerged since the development of Giza++ (e.g., Marian, xnmt, OpenNMT, Sockeye, T2T[21]); (months 7-12).
3. Assessment of alignment methods with other language pairs as available (work has already been done on classical automatic alignments with IBM models and HMM between Ancient Greek and Chinese, French, Italian, Modern Greek, Persian and Spanish); (months 13-24).
4. Integration of training data, including human-curated/produced alignments as well as machine readable dictionaries, language specific versions of WordNet (e.g., development of WordNets that exist for Greek and Latin); (months 25-36).

(3) Crane will be responsible for the development of content to exploit and drive development:
1. Development of a curated bilingual edition of the *Iliad* and *Odyssey*, which includes automatic alignments (months 1-6), the morpho-syntactic analyses from the Perseus Treebank, machine readable lexica, and as many openly licensed resources as possible (months 7-12), as well as curated alignments for selections of the two epics (months 24-36).
2. Integration of translation alignment into all the courses that he teaches, including an advanced CS course on working with corpora, introductory and advanced Greek, and classes on Greek literature in English translation (months 1-36,, with ongoing blog posts to report on progress).
3. Supervision of students to develop automatic and curated alignment projects. These include (1) addition of morpho-syntactic information to the existing English curated alignments of the Divan of the Persian poet Hafez (months 1-12), (2) French, Latin, and English versions of Bodin's early modern Six Books on the Republic (months 1-12); (3) new projects (months 13-36)
Risks and risk mitigation: There is always a risk that software development will take longer than anticipated but here we are building on prior work that has identified methods that scale on the backend and developed front-ends to support key functions (such as translation alignment). Sklaviadis' work includes publishing evaluations of the quality of the automatic alignments. The biggest challenge to Crane's work lies in availability of students with suitable skills but we can bring students in from other schools during the summer.
Assessment: (1) Have we been able to add the planned services with adequate speed/performance? do third-parties build on our software, data or more general methods? How much direct usage do the translation services attract within the new Perseus?

Final Products and dissemination

Our goal is to support emerging voice-to-text services so that our audience can use speech to pose queries and so that they can hear English translations and (insofar as possible) ancient languages. In general we follow the W3C guidelines for Web Accessibility (https://www.w3.org/standards/webdesign/accessibility). **Products**: (1) alignments published as data on GitHub; (2) software on GitHub. Final Product+Dissemination: (1) a new Perseus featuring the funded work; **Dissemination**: publications, talks, integration into ongoing workshops in Germany as well as at Tufts.

---

[21] https://marian-nmt.github.io/; https://github.com/neulab/xnmt; http://opennmt.net/; https://github.com/awslabs/sockeye; https://github.com/tensorflow/tensor2tensor.