Faculty Development Plan for █████████

Updated 15 August 2022

Personal Assessment

My current strengths include the following:

1. Knowledge of the Machine Translation (MT) field and involvement in the professional MT community: I have actively participated in the evolution of MT technology for the past 47 years. I have been a member of and served in leadership positions in various professional organizations during this period, including ACL, AMTA, TAUS, and IAMT. I have been the president of AMTA for the past 4 years and IAMT for 2 years. I have also attended and presented at various conferences, including ACL, NAACL, COLING, AMTA, MT Summit, LREC, and TAUS.
2. Industry Experience in research, development, and production: I have worked at BYU, IBM, Microsoft, and the Church of Jesus Christ in both R&D and production positions, and I have broad experience across the spectrum of MT and NLP research and application.
3. Close contact and relationship with the heads of MT groups at major companies: I personally know senior people and heads of MT groups at many companies, including Google, Meta, Microsoft, Amazon, Systran, Unbabel, RWS, Lilt, and others. These connections have enabled me to facilitate job opportunities my students and interns.
4. Passion for languages, computing, and MT: I speak Portuguese fluently and Spanish to some degree, and I have interacted regularly with translators for dozens of languages and computer scientists around the world. MT and NLP have been my passion for almost 5 decades.
5. Intense desire to break down all language barriers to take the gospel of Jesus Christ to the world: my experiences as a priesthood leader (esp. mission president) and working for the Church for 10 years have infused me with a desire to use my knowledge and skills to spread the gospel to the world. I share this desire with my students regularly.
6. Extensive experience teaching and working with young adults, especially returned missionaries: I have lead, taught, and supported young adults and RMs for over 20 years.
7. Dedicated to helping students succeed in their families and careers and in making significant contributions to the world and the Church: my work with RMs and YSAs has engendered the tremendous love I have for them and desire to help them succeed in life and in the gospel. I incorporate gospel teachings naturally whenever I can, and my focus on languages and MT is extremely compatible in that regard.

Improvement Opportunities

1. Course organization: although my knowledge and experience are extensive, organizing content for the MT course I have taught has been challenging. The Effective Teaching Seminar was most helpful in creating a comprehensive syllabus. I will use this training for my other courses as well.
2. Programming skills: While I was a very proficient programmer earlier in my career, my skills are somewhat dated. I am focused on rapidly coming up to speed in Python and newer programming tools and methods developed in the past two decades.
3. Involving students during classes: My first MT class in Winter 2021 was very lecture-oriented, but after receiving feedback and instruction from various sources, I think I did much better at involving

students when I taught it again in Winter 2022. The Effective Teaching Seminar has provided additional Ideas I will implement this fall and beyond.

4. <u>Administrative aspects of grant writing, purchasing, hiring</u>: I have never worked in academia, and I have not had to deal with writing grants or purchasing my own equipment. I've had experience hiring and managing many people, of course, but the academic environment has its own idiosyncrasies in this regard. Learning these administrative aspects will come with time, I'm sure.

5. <u>Return to regularly publishing and attending research conferences after an absence of years</u>: for the past 13 years serving and working for the Church, I have been principally involved in the implementation and use of MT systems and other translation tools, but I have not been involved in the research and technology side of the field, although I have strived to keep up to date on all the latest advances. It is fun and refreshing to return to that, but it will take some time and effort to fully engage.

My goals to improve in many of the above areas are included in the Goals and Status sections below.

Teaching Goals and Status

1. Revise and improve the CS 401R course I have just begun teaching on MT. I have incorporated the feedback from previous students and from ███████ on Winter 2021 class I taught as an adjunct, as well as new content based on the rapidly advancing MT field. I will yet incorporate feedback from students who attended the course in Winter 2022.

2. Create a curriculum and prepare for an advanced graduate course on MT, CS 601R, that I will teach next winter semester.

3. Organize, prepare for, and teach two sections of CS 111 next winter. I will significantly improve my Python familiarity and skills by attending CS 111 lectures taught by my CS colleagues throughout the fall semester.

4. Teach my CS 401R course on MT again in the fall, continuing to update and improve it, incorporating knowledge gained from the NFS Spring Seminar and the Effective Teaching Seminar I attended into the course, especially regarding a more comprehensive and descriptive syllabus. I already have a draft of this syllabus. I will also apply for the course to have its own course number.

5. Mentor the students in the research lab I am establishing (see goal (3.) in the scholarship section below). I have already been meeting with and mentoring 2 undergrads in my lab, one of whom will be the TA for my MT class in the fall.

Scholarship Goals and Status

1. Identify and pursue the research for and publication of two-three papers. Given the May-June 2022 submission deadlines for EMNLP, COLING, and AMTA, it was unlikely that research and corresponding papers could be ready in time for this year. However, through collaboration with a former intern I mentored at the Church who is now in the CS MS program at CMU, as well as with a former student in the MT class I taught in Winter 2021, we performed additional experiments and wrote and submitted a paper about an automatic interpretation and video dubbing system (Lingua), which they participated in creating and enhancing. The paper was accepted for presentation at the AMTA 2022 conference in September. Additionally, another former student is working on a paper based on the project on weakly-supervised MT he did for my most recent MT class, for which I will

be co-author, and we intend to submit it to the WMT conference (part of EMNLP) held in December. These two papers already minimally satisfy this goal, but more possibilities exist this fall.

2. Current research plans (which may evolve) include:
    a. Creation of a large multilingual corpus for MT and other NLP-related research based on the Church's translation databases. While creating this corpus should be very doable, making it publicly available may delay and be challenging, depending on the Church's IP department. I have already acquired all the data and it is currently being cleaned and prepared in the lab mentioned in goal (3.) below.
    b. Extending work done on "complete" Multilingual NMT (cMNMT), published (by Google) in late 2020 using the corpus created in point (a.) above. cMNMT depends on training with segments that are multi-way-aligned across multiple languages. Much of the Church's data exhibits exactly this property.
    c. Extension and refinement of the English-ASL MT system created by a student in my Winter 2021 MT class, which could have been published if we'd had more time (and the student didn't graduate!).
    d. Extension of the work on automatic interpretation and video dubbing described in the paper to be presented at AMTA 2022, which was mentioned in goal (1.) above.

3. Create a research lab that includes at least 1-2 grad students and 2-3 undergrads by the end of the year. I already have 2 undergrads working and 1 grad student committed for fall. The lab will focus on the topics listed above, and very likely will include other low-resource language MT research, since many students in my MT class have been very interested in this topic, and I am entering into various collaborations with external colleagues on the same topic (see goal (5.) below).

4. Establish a monthly brown bag cross-departmental group on MT- and NLP-related topics. I've already spoken with faculty in Linguistics, Digital Humanities, and CS about this and I will attempt to start it this fall, but I have already observed the challenge in doing so, as many faculty seem to be very focused on their own work.

5. Explore potential collaborations with my various AMTA contacts from MT groups at companies and universities, and establish a collaboration with one of these by the end of the year. As a result of attending NAACL in July, I already have 4 promising possibilities, including from Microsoft, University of Washington, New York University, and Johns Hopkins University. I will likely have to pick just a couple of these, or somehow combine the work involving them so that I am not spread too thin.

6. Identify and discuss grant opportunities with contacts and with CS faculty colleagues. Write and submit (possibly with a co-author) a grant application. The NFS Spring Seminar and the Grant-Writing Workshop I attended in May have both provided excellent guidance on this. I am already in discussion with Microsoft about a grant to perform the multilingual corpus described in point (2.a.) above, and another grant I may become involved in is with my University of Washington contact, who is submitting a grant to NSF to develop language technologies for crisis response situations.

Citizenship Goals and Status

1. Continue as AMTA president and as general chair of the upcoming AMTA 2022 conference in September. This is a significant undertaking and is serving as my main citizenship project. At the end of year, I will become a "counselor" on the AMTA board as another president takes over, but I will still be very involved in the organization, as I have been for the past decades.

2. As a member of the external relations committee, assist in creating the first version of a partner engagement strategy for our CS department, in managing and updating the Capstone and Partnering pages on our website, and in planning for our Donor board in the TMCB.
3. Serve as reviewer for submissions to AMTA 2022 (completed), the Language Resources and Evaluation Journal special issue on Translation Platforms, and possibly other conferences or peer-reviewed publications.
4. Invite 2-3 top scholars and practitioners in my field of Machine Translation (MT) and Computational Linguistics (CL) to speak at the weekly CS Distinguished Speaker seminar during the coming school year. I have already invited two speakers – the head of MT at Microsoft, and the VP of Language Technology at Unbabel (and former head of MT at Amazon) – and they have both accepted. One will speak in November and the other next April. I also have a very promising third candidate who is a highly respected professor of CL at Johns Hopkins University and ACL Fellow.
5. Invite faculty from other departments to participate in monthly brown-bag discussions of our current work and other relevant papers in our areas of intersecting interest. This is the same as (or very similar to) goal (4.) in the scholarship section above.
6. Actively seek to involve students in my MT courses, lab, and research from diverse backgrounds and underserved groups. Because of my focus on MT for a huge number of languages, and mainly on MT for "low-resource" languages, I expect to involve: returned missionaries, including many women, who have served in countries around the world and who speak a wide variety of second languages, students of all ethnic, racial, and cultural backgrounds who have native languages other than English, and students of all orientations who, because of the very diverse backgrounds of other students in my courses and lab, will hopefully feel that they fit in and are as respected and valued as anyone else. My main criterion for inclusion is a passion for language and computing, and I hope that students will share my additional passion, which I speak of often, to break down all language barriers so that the gospel of Jesus Christ can go to all the world.

Relationship between my goals and department and university aspirations and needs

I am grateful that the focus of my teaching and research fits so well with the need to be inclusive of people from all backgrounds and with the mission of the Church to take the gospel to every nation, kindred, tongue, and people. In his Second Century address, President Kimball spoke of the need for BYU to become "the acknowledged language capital of the world" and he said the Church "look<ed> forward to developments in <BYU's> computer-assisted translation projects." That talk was given in the first year I began working on machine translation! All my goals target the realization of this prophetic vision. Furthermore, while the first reason I came to BYU was to fulfill exactly this vision, the second and equally important reason was to help raise up a new generation of young people who were able and dedicated to spreading the gospel to the world through the use of this technology. Hence, I have goals to ensure that my teaching and mentoring accomplish this purpose most effectively.