Linguistics and Language Technology: A Practical Appraisal

Mohd Asjad Husain

Ph.D. Scholar, Aligarh Muslim University, Uttar Pradesh, India

ABSTRACT

In developing nations, Computational Linguistics (CL) and Language Technology (LT) are areas of wide research possibilities in linguistics, language teaching and silicon industry. These intersections and their combinations offer countless professional opportunities for practising as well as prospective researchers in the field of linguistics and language technology. Effective researches in these areas require interdisciplinary expertise in linguistics and Information Technology (IT). However, in the context of developing countries, researchers face countless challenges regarding their competence and performance while conducting practical research. This paper seeks to understand the ground level realities of scholarly expectation from Language Technology (LT) researchers. The paper will provide an insight into their professional lives, previous academic qualification/experiences, and facilities they are provided with. Moreover, it will help in gauging a practical understanding for creating better and more realistic task force for effective researches.

Keywords: *language technology, computational linguistics, researcher, research, competence, performance*

A. INTRODUCTION

Virtual or real, language is a central component to any kind of human interaction. From interpersonal to mass communication, irrespective of ever-changing modes and methods, language is essential for the transmission of basic information as well as for the intellectual development of its users. From printed books, newspapers, magazines, radio, television, computers to the latest smart phones and other highly sensitive personalized gadgets in the present century,

ways and modes of communication keep changing as per popular culture(s) of the time. Language has been flexibly compatible to fit in and work with all types of media according to their technical requirements. For this, linguists have worked meticulously with the experts of related media fields.

In the recent past, advancements in the technical industry have attracted experts on technical fields to undertake studies in linguistics to make Information and Communication Technology (ICT) a part of human life. Consequently, for students and scholars in linguistics, Computational Linguistics (CL) has emerged as a growing career option with multidimensional job profiles from small firms to the multinational giants in the silicon industry. This paper seeks to explore basic understanding of CL from the point of view of long-term career goals and its expectations among the students and scholars in the field of linguistics. It sees the entire issue from the perspective of linguistics (not from computer science).

Traditionally, operations in CL were performed by computer scientists with specialized knowledge of applying computer technology to process natural languages. Soon,

A general lack of expertise in linguistics was realized by computer scientists. As a result, this realization opened up the field for people with strong background in linguistics. This made computational linguistics a regular part of linguistic studies around the globe. And in the present time:

> Computational linguists often work as members of interdisciplinary teams, which can include regular linguists, experts in the target language, and computer scientists… draw upon the involvement of linguists, computer

scientists, experts in artificial intelligence, mathematicians, logicians, philosophers, cognitive scientists, cognitive psychologists, psycholinguists, anthropologists and neuroscientists, among others. ("Computational Linguistics", 2017)

As a result, as Gallimore (2013) says, CL has become "an emerging industry, where difficult science is applied to very difficult problems with often mixed results". These (very difficult) problems call for multidisciplinary expertise. On one hand, computer scientists are expected to be competent in linguistics, on the other, linguists are expected to develop required technical skills, most importantly, a fresh technical perspective to view their own field. Gallimore considers it as a crucial concern which:

> Touches upon 2 real issues that faces computationally-minded linguistics graduates in particular: First, the vast majority of work in Computational Linguistics is now happening in software development environments, and software companies tend to value Computer Science degrees highly. Secondly, the way that the technology world approaches linguistic analysis is predominantly statistical, and linguists tend to be taught rule-based approaches instead. In fact, some would argue that Computer Scientists and Linguists are a long way apart in how they view language, and that mathematicians are better qualified to work in Computational Linguistics than most linguists. (2013)

Gallimore's strong comment calls students and scholars in linguistics to understand the practical requirements of being in the field. According to him, attaining theoretical knowledge, familiarizing oneself with the technical terminology of the field, and earning degrees are not sufficient. He offers following suggestions:

1. Do the Masters:
Gallimore sees a master degree with specialization in computational linguistics as the starting point. He holds that without a master degree "it's not possible for a Linguistics graduate to be taken seriously for the majority of positions in Computational Linguistics" (Gallimore, 2013).

2. Learn to code:
Practical understanding of programming and coding is central to be effective computational linguist for in the practical field, 'computational' (knowledge of computer languages) is as important as 'linguistics' (and its subfields). To be expert in one and know nothing about the other creates big challenges in the professional life. He further adds "most universities will welcome Linguistics grads with limited or no coding experience, [but] most companies developing language technology won't" (2013).

3. Get some experience or build something:

Practical Exposure and experience is always desirable and it is something that universities (project-in-charges) and companies value. Gallimore, a recruiter in computational linguistics, says:

> That sounds scary but actually it's quite simple – pretty much all of you [students and scholars in linguistics] will build something interesting on your Masters, but most of you won't tell anyone about it. If Internships aren't for you (and I get that), then why not take your Masters project, develop it a bit in your own time, and then get it on-line. Built a PoS tagger? Great! Can I see it? Developed a prototype ASR? Awesome. Put a nice UI on it, get it hosted somewhere and tell us about it on your profile. You won't believe how powerful this is. (2013)

These are some important suggestions from Gallimore – a classically-trained linguist in Computational Linguistics and Natural Language Processing. He has nearly 10 years' experience in recruiting technology professionals for software companies.

B. Required skills to be in the field of CL

Computational Linguistics is broadly divided into two categories; (i) Theoretical Computational Linguistics and (ii) Applied Computational Linguistics:

i. It discusses issues in theoretical aspects of linguistics and also deals with cognitive science. It talks about formal theories that provide basis for linguistic knowledge which people require for producing, understanding and analysing language. Linguists working in this branch of computational linguistics develop functional programs simulating features of the human language faculty and implement them as computer programmes.

ii.   It emphasises the application-based outcome of modifying human language. Tools, techniques, methods and applications in this branch are often characterized as (human) language technology or language engineering. It focuses on the creation of software products with human-like language data to enhance the acceptance of the software and the productivity of its users. ("Linguistics Jobs: Computational Linguist", n.d.)

As CL is a challenging technical field of work, to be able to work in both the categories and develop intelligent machines which deal with natural language, students and scholars from linguistics need a good command over both: programming and theories of linguistics. CL creates systems that can complete artificial intelligence based complex tasks like speech recognition (Siri, Cortana), machine translation (Google Translate), speech synthesis, grammar checking, spelling checking and many other applications for integrating speech recognition with language processing techniques. They are usually able to solve problems in the following areas: Natural Language Interfaces, Machine Translation, Grammar and style checking, information retrieval, Document processing and Computer-Assisted Language Learning (CALL).

Familiarity with Natural Language Processing (NLP) applications and experience in programming languages is essential. Although, there are hundreds of computer languages in existence, on the basis of above discussion, following are important ones for conducting effective research and for being an effective computational linguist:

1. C++
It is a general-purpose programming language. It has imperative, object-oriented and generic programming features, while also providing facilities for low-level memory manipulation.

2. Java
It is also a general-purpose computer programming language that is concurrent, class-based, and object-oriented.

3. Python
It is another important and widely used high-level programming language for general-purpose programming.

4. PEARL (Process and Experiment Automation Realtime Language)
It is a computer programming language designed for multitasking and real-time programming. Apart from these, BETA, Milk, PRAAT, ELAN, AUDICITY and FLEX are also used in the field.

5. Skills in Linguistics: Despite above mentioned technical skills, computational linguists are also required to have in-depth knowledge of linguistics which includes the understanding of phonetics, phonology, morphology, syntax, semantics and discourse. Also, students and scholars in CL are advised to be fluent in their mother tongue and multiple foreign languages (Gallimore, 2013; "Become a Computational Linguist", n.d.).

C. Problems and Challenges: Researchers in CL

In order to understand the problems and challenges faced by the researchers in CL, the researcher conducted semi-formal unstructured interviews with the scholars and the researchers in the field of CL. Some broad and open-ended questions were prepared by the researcher in advance to maintain the direction of the conversation. Participants were asked to be flexible and were allowed to express their opinion without following any particular pattern of answer. Following questions were prepared:

i.   Does your university/college teach CL? When does it start teaching it?

ii.   Does your university/college have trained and competent faculty members for teaching CL?

iii.   Do students/scholars have significant background in the CL?

iv.   Is there any gap between what students and scholars study (BA and MA syllabi) and professional requirements as a researcher in CL?

v.   Are they aware of the technical requirements for being competent researchers in CL? For example, computer programming languages and other related software.

vi.   Are they aware of basic technical skills like JAVA, C++ and others?

vii.   Can they code/program computer languages and develop software?

viii.   When and how they decide to be a computational linguist?

ix.   What are their motivations to choose CL as their career? Do they choose it due to their interest or due to some other reasons?

x. Is your university/college infrastructurally prepared to teach CL?
xi. Do you have a computer lab in your university/college?
xii. How often do you practice programming/coding in the lab?
xiii. Do you learn programming/coding during the course of study?
xiv. Does it provide required administrative support? For example, arranging industrial tours, internships and conducting workshops, lectures etc.
xv. How often students in CL get part time/full time employment: internships, projects, teaching, industrial jobs, etc.
xvi. Do the faculty members encourage students and scholars in the CL?

Questions were based on, but not limited to this list. Other issues were also incorporated on the basis of the responses from the interviewees.

D. Findings

Based on interviews conducted for this paper, the researcher found following problems:

1. Majority of universities/colleges do not teach CL. It is taught only in few universities/colleges.
2. Even the universities/colleges which are teaching it, do not have trained and competent faculty members. It is found that faculty members who are trained on traditional linguistics and do not have any background in technology are teaching CL. (some of them cannot even work on the computer).
3. Some universities/colleges have stopped offering courses/papers in CL in the want of faculty members.
4. Even the universities/colleges which are teaching CL include it only in their MA level syllabi. No or least exposure is provided at the bachelor's level.
5. Papers in CL focus only on the theoretical aspects. No or least practical exposure is provided.
6. Majority of students/scholars do not have any prior background in CL and computer programming languages. (Some of them do not even have their own machines).
7. There is a wide gap between the CL syllabi and required skillset in the professional field. (What students/scholars study is sometimes not even required).

8. Though most of the students/scholars are aware of the technical requirements for being competent researchers in CL; such as, computer programming languages like C++ and Java and other related software like PRAAT, ELAN, AUDICITY and FLEX, they know names only and cannot use them.
9. Apart from some exceptions, majority of students/scholars cannot code/program.
10. Students/scholars often decide to study CL only on the basis of secondary information or peer discussions. They are often attracted towards CL for *big salary packages and glamour* it offers (if it does). It is found that they do not make a very thoughtful decision.
11. Majority of students/scholars reported that their university/college is not prepared infrastructurally to teach CL. Some of the universities/colleges do not even have a working language technology lab. Some of them have it, but it is not functional. Some general problems are as follows: a. Computers do not work.
    a. They do not have working internet connection.
    b. They do not have required software.
    c. They do not have working antivirus software.
    d. Labs have basic computers which do not match the configurations required for CL.
12. Universities/colleges do not provide skilled faculty members or technical support In charges to teach coding/computer programming to students/scholars.
13. Students/scholars do not have any practical understanding and/or experience of coding and computer programming.
14. Sometimes, students try to learn technical skills by themselves either by a. Contacting computer science departments Attending private courses by themselves.
15. It is found that universities/colleges rarely provide any administrative support to such students/scholars.
16. Universities/colleges rarely arrange company tours, internships, and conduct workshops and lectures to make their students/scholars aware of the developments in the field of CL.
17. Employment ratio of students/scholars in CL is significantly low.
18. It is found that university/college faculties motivate students/scholars to study CL, but rarely provide practical support.

E. SUGGESTIONS

i. Courses/papers in CL should be introduced and included at all levels i.e. bachelors, masters and Ph.D. Moreover, at the-
   a. Graduation level: at least a complete unit should be included in the syllabus.
   b. Post-graduation level: Paper in CL should include practical understanding of coding/programming software. These papers should be taught by the trained faculty members. Product based technical projects should also be made part of curriculum at this level.
   c. Ph.D. Level: Universities/colleges/departments should collaborate with other institutions and government bodies who are actively engaged in the conducting research projects in CL.
ii. Universities/colleges should form intra-departmental study groups of CL enthusiasts.
iii. Universities/colleges/departments should include CL in their policies and should hire theoretically competent and practically trained faculty members with strong background in technology.
iv. Those universities/colleges which have stopped offering courses/papers in CL should address the reasons of its discontinuation and start afresh.
v. Courses/papers should be revised as per the market needs. Both theoretical and practical aspects should be given appropriate space in the syllabi.
vi. IT experts should be hired to teach programming languages and other important software to develop a strong technical base in students/scholars in CL.
vii. Expert lectures and workshops should be organized to make students aware of the opportunities in the field of CL.
viii. Universities/colleges should improve infrastructurally to facilitate the teaching of CL. Lab maintenance committees should be constituted to make sure that:
   a. Computers are working properly.
   b. They have working internet connection.
   c. They have required software.
   d. They are updated on the required configurations.
   e. They have working antivirus software.
ix. Universities/colleges should have administrative collaborations with their computer science departments to provide required help to the students/scholars in CL.
x. Universities/colleges should arrange company tours, internships and conduct workshops for students/scholars in CL.
xi. Universities/colleges should develop collaborations with software companies and government bodies to enhance the employability of the students/scholars.
xii. Universities/colleges should conduct awareness lectures for their faculty members, thereby, enabling them to be effective mentors.

F. CONCLUSION

General competence in applied linguistics as well as NLP and other related software is the basic requirement for researchers to be able to take long-term and productive jobs in the blooming silicon industry. But many of the aspirants and educational institutions are lacking behind due to lack of information and/or inconsiderate policies. Thus, this paper provides an in-depth understanding on issues and challenges faced by students, scholars as well as institutions in developing nations. It offers an open-ended blue print for market based steady professional development in the filed.

REFERENCES

1. Gallimore, N. (2013, August 21). *From Linguistics to Computational Linguistics.* Retrieved From URL: https://nlppeople.com/from-linguistics-to-computational-linguistics/
2. Computational Linguistics. (2017, December 27). In *Wikipedia*. Retrieved From URL: http://blog.apastyle.org/apastyle/2009/10/how-to-cite-wikipedia-in-apa-style.html
3. Become a Computational Linguist: Career Roadmap. (n.d.). www.study.com. Retrieved from: https://study.com/articles/Become_a_Computational_Linguist_Career_Road map.html
4. Halff, H. (1986). Instructional Applications of Artificial Intelligence. Retrieved from: https://www.researchgate.net/publication/23466789 9
5. McCulloch, G. (2017, September). Linguistics jobs - interview with a Text Analyst [Blog post]. Retrieved from: http://allthingslinguistic.com/search/computational+
6. McCulloch, G. (2017, December). Linguistics jobs - Interview with a Linguistic Project Manager

at a Language Tech Company [Blog post]. Retrieved from: http://allthingslinguistic.com/search/computational+

7. Eisner, J. (2016, August 21). How is computational linguistics different from natural language processing? Retrieved from: https://www.quora.com/How-is-computational-linguistics-different-from-natural-language-processing

8. Computational Linguistics. (2017, December 25). In Wikipedia. Retrieved from: https://en.wikipedia.org/wiki/Computational_linguistics

Author's note:

Mohd Asjad Husain is a former Fulbright Foreign Language Teaching Assistant (FLTA) at the University of California, Davis and Cultural Ambassador of India in the USA (2015-2016). He is a Ph.D. scholar in the Department of English, Aligarh Muslim University, India. He is currently working on 'English Language Teaching in Postmodern Era: Principle and Practice'. His areas of interest are: decolonization and ELT, use of technology in second/foreign language learning, second language learning, postmodernism in ELT, autonomy in language teaching, critical pedagogy and constructivism.