

Teaching Computational Linguistics in English Language Academic Programs

Essam Hassan Al-Mizgagi

University of Science and Technology, Sanaa, Yemen

Email : esamhasan10@yahoo.com

Orcid  : [0000-0002-2384-9272](https://orcid.org/0000-0002-2384-9272)

Received	Accepted	Published
14/7/2024	27/10/2024	31/10/2024

 : 10.5281/zenodo.14031289

Cite this article as : Al-Mizgagi, E. H. (2024). Teaching Computational Linguistics in English Language Academic Programs. *Arabic Journal for Translation Studies*, 3(9), 186-204.

Abstract

This study aimed at suggesting a syllabus named “Introduction to Computational Linguistics” to be taught within the programs of Language Teaching and Translation in the Yemeni Universities. This course could be a threshold for scientific study of language through dehumanizing the routinely-repeated processing, analyzing and even production of natural human languages to save time and effort, and avoid bias and subjectivity. The study method relied on a review for several similar programs from well-known universities which included this course within their study syllabi. It also used a field survey through which academics and professionals who gave their insights about the feasibility of including CL within the programs of language departments in the Yemeni Universities. A review to previous research papers and books highlighted the importance of teaching language processing and programming at language programs as well that highlight the importance of giving training to students of Language programs to introduce students to Natural Language Processing (NLP) skills such as tagging, tokenizing, parsing, sentiment analyzing, machine translation, text production, analysis and summarization that could be handled in later courses. The study concluded with a strong recommendation for tradition language schools to include CL course within their academic language programs either within the BA and/or MA program syllabi either as ready-made software track or as a programming track. Later appraisal might give insightful feedback for the development phase to the concerned Language program.

Keywords: Digital Humanities, Python, Computational Linguistics, Natural Language Processing, Translation

Introduction

The Recent advent of Artificial Intelligence (AI), and the emergence of Digital Humanities (DH) has, recently, become a significant trend to understanding the realm of humanities in general and language programs in particular more scientifically and objectively. Computational Linguistics (CL) is the scientific study of language processing from a computational perspective which can be deemed as . It is a field that combines linguistics and computer science to develop computational models of language and to create new applications for language processing. (Jurafsky & Martin, 2020). One of the primary functions and mechanisms of CL is to quantify the unquantifiable nature of human languages. It is an intrinsic part of Digital Humanities (DH) that tries to computerize the qualitative and subjective nature of human sciences (Sakthi Vel, 2017; Voutilanen, 2005). Language programs, such as Translation, ELT, Applied Linguistics, and Literature, are directly concerned with Computational Linguistics (CL) due to the theoretical nature of the course. CL more specifically, takes the linguistic role in this respect since it represents a serious departure from the qualitative subjective nature of appreciation of the human language for a quantitative objective mechanism of thinking of this realm of knowledge through techniques such as quantification, annotation and attribution. Thanks to CL, apparently unquantifiable emotions and attitudes can be calculated and measured to bring about concrete evidence objectively through recognizing, measuring and quantifying linguistic and metalinguistic tokens such as connotation and expressivity. This quantitative turn, with statistical procedures, is gaining in popularity in many subfields, from usage-based morphology to Cognitive Semantics, and from phonology to discourse analysis (Levshina, 2015).

Computational Linguistics (CL) can be defined as a science that seeks to develop the computational machinery needed for an agent to exhibit various forms of linguistic behavior. By “agent,” we mean both human beings and artificial agents such as computer programs. By “machinery,” we mean computer programs as well as the linguistic knowledge that they contain (Fasold & Connor-Linton, 2013).

Computational Linguistics in its broad sense has been for a long time since human being depended on calculating and analyzing their given data using primitive tools such as stone tools and, later, tallies on paper (Britannica, 2022). In its narrow sense which refers to the use of computer, Computational Linguistics started its development in the 1950s driven by the practical need to create the systems for machine translation (Mitkov, 2009). It is the computational linguistics that is driving the developments which we conceive to be the “artificial intelligence” (Ivashkevych, 2019). Since the 1990s, computers have become steadily faster and have provided access to increasing quantities of on-line linguistic data (the Web being a prime example) (Fasold & Connor-Linton, 2013). Methods based on statistical analyses of such data have dramatically improved the accuracy with which systems carry out tasks like understanding the syntactic structure of a sentence. The

success of such methods has raised questions about how language is represented and processed by the human mind, and particularly about the role of statistics in language understanding. It also suggests that humans might learn from experience by means of induction using statistical regularities. (Fasold & Connor-Linton, 2013).

With respect to the official status of Yemen as a third world country, Yemen has a bureaucratic educational system that is not elastic enough to cope with potential changes for the purpose of developing and updating its educational system and infrastructure in terms of pedagogical requirement such as syllabus development especially during the civil war that Yemen is witnessing these days (Muthanna & Karaman, 2014). English programs are no exception because they still depend on the traditional scope of objectives that does not exceed to technical text processing. Basically, high school students, on the other hand, do not have sufficient knowledge and use of information and communication technology (ICT) competence. The close future of these is mainly at the higher education institutions which is required to include ICT competence as intrinsic part of their professional skills to yield an enquiry-based education (Kubrický & Částková, 2014). According the National Strategy for the Development of Higher Education in Yemen, *the Government's Higher Education Project plans the creation of the Yemen Foundation for Information Technology to lead the development and use of ICT in Yemen's universities and colleges* (NATIONAL STRATEGY FOR THE DEVELOPMENT OF HIGHER EDUCATION IN YEMEN, 2005). The strategy mentioned general initiatives in this regard though.

The Study Rationale

Due to the proliferation of the computerization of Human languages and language programs, linguists can take part in building a sound language computerization in the fields of Translation, Applied Linguistics, Literature, and ELT. Linguists, therefore, may learn utilizing language-processing software and programing so that they can be present with a solid linguistic knowledge and take part when programmers work on language-centered processing and annotation such as; libraries for tokenization, tagging, segmentation, parsing, Translation, even more concise sentiment analyses, and AI utilities like LLMs. This study, therefore, suggests introducing two-track modules that aim at equipping students with language processing skills through computational tools; the first track is introducing students to ready-made software and website services, and the second track is qualifying students with programing abilities through learning from linguist-oriented materials like the textbook suggested which is "Python for Linguists" (Hammond, 2020) for students of English Pogroms 3 which can be a foundation stone for the career that majors in computational linguistics.

Statement of the study problem

Due to the advent of technology-based language processing services such as machine translation and AI, websites, software and applications, Language programs, up to now, don't have a fair role and should take a due part in the process of improving the computation of natural languages through including CL within their academic study programs due to their deep theoretical knowledge about the natural languages and their mechanisms (Kim, Y. et al, 2020; Jones, C. et al., 2018; Smith, A. et al, 2016), linguists should be an intrinsic part of the process of language computerization.

The study objectives

The goal of a research proposal is twofold:

1. to present and justify the need to include a course named "Introduction to Computational Linguistics" within the study plans of English language programs in Yemen, and
2. to propose a course framework according to which these programs can follow to initiate teaching "Introduction to Computational Linguistics".

The significance of the study

This study tries to give a strong recommendation to include CL as an introductory course within the English programs being taught in Yemen and other third world countries that are in similar conditions. Johnson (2011) claims that there are two basic parts; scientific that should be handled by linguists and technological that is missioned to programmers who develop tools to handle Natural Language Processing (NLP). Programmers cannot deal with languages based on their shallow knowledge of natural languages. Programmers, therefore, should seek collaboration with linguists who have deep knowledge of languages so that they can handle the linguistic nuances and delicacies more professionally through incorporating Computational Linguistics into language academic programs (Jones, C. et al., 2018; Kim, Y. et al, 2020). Johnson (2011) continues emphasizing the importance of the scientific suggesting it is so deeply rooted that needs to have linguists who are familiar with fine-tuned problems and solutions for complicated linguistic issues such as descriptive grammar found in e.g., Baker (1995), Huddleston and Pullum (2002) and McCawley (1988). Then he concludes that there is no reason to expect an engineering solution to utilize all the scientific knowledge of a related field (Johnson, 2011). This can be regarded as the threshold to embarking the newly adopted approach of 'Digital Humanities'. More specifically, this proposal targets English-related programs such as; translation, applied linguistics, arts and ELT.

As a matter of fact, there is no need to say that, these days, more and more sciences are being computerized and automatized. Linguistics and translation are no exception. The

proliferation of a handsome number of language-centered software and applications in different operating systems, such as IOS, android and windows, is remarkable. The course of Computational Linguistics is meant to equip students with basic introduction to language-related skills such as tagging, tokenization, parsing, semantic analyzing, machine translation, and text production, analysis and summarization. Machine Translation (MT) and Language Models (LM), for instance, are used worldwide. Based on these this course, the later courses in module, say NLP, will give sound practice to these skills. With respect to linguistics, the linguistic defects that can be notice in MT is a good example for the necessity to learn program, as a product, to fix language-based errors (Jones, C. et al., 2018). Hammond (2020) declared that programming is a useful skill in many areas of linguistics and translation and in other language-related fields like speech and hearing sciences, psychology, psycholinguistics, quantitative literary studies and other digital humanities programs. Within linguistics, it used to be the case that programming skills were required only for computational linguists, but this is far from true these days. Programming now is used in phonology, syntax, morphology, semantics, pragmatics, psycholinguistics, phonetics, discourse analysis, essentially every area of linguistic investigation (Hammond, 2020). Computational Linguistics has become the basis for solving many practical tasks in the language industry. When providing the students of linguistics and translation with the basics and the methods of the computational linguistics, they will cope with the current view for human language that can be handled and utilized through these gadgets that, inevitably, are take over the profession. The students, then, need to widen their views on linguistics and show computational linguistics as a perspective field for their possible future to be engaged with it. it is worth introducing the basics of computational linguistics (CL) in general and programming that can help students start with natural language processing (NLP) in particular to the students of linguistics and translation.

The study method

To achieve the first objective of this study, a mapping review to the English programs which included CL withing their syllabi was carried out to yield qualitative description of some overseas academic programs in which the proposed study course was taught within any language academic program in the faculties of Humanities, Languages, and Communication that had this course/module in their Program Specification Documents PSD. Another field survey was conducted through distributing a questionnaire to stakeholders (academicicians, researchers, under/post graduates) to get acquainted to what extent CL was feasible to be taught in the programs of English Department. Both the review and the survey can help reach a commonsense amongst Yemeni academicicians about the feasibility of including CL within their academic programs

With respect to achieving the second objective of this study, a proposal was suggested based on two options: the first option is product-oriented in which English language

programs may depend on ready-made software and website services to equip students with abilities and skills that enable them to handle language processing such as text summarizing, tagging, translating, and token quantification; the second option is process-oriented which aims at equipping students with a programming language such as Python to handle the process of text processing from the scratch. For this track, Michael Hammond's textbook 'Python for Linguists' published by Cambridge University Press, 2020 was proposed. The rationale for choosing this Textbook is because Hammond is, basically, a linguist and works for Linguistics Department at the University of Arizona. Hammond's remarkable contributions are focused on computational linguistics (CL) and Natural Language Processing (NLP) in other languages such as; Java and Perl. In a review for this book in a MIT Press Direct, Roth and Wiegand declared that, this book targets to linguists with no prior programming background. At the end of the review the reviewers declared that this book *would make the book a great companion for a foundational programming course targeted at linguists* (Roth & Wiegand, 2021). Besides, this book is one the most recent textbooks in this area up to the date of this report.

It is worth to say that the first option is a shortcut to the study course since it, directly, provides students the required skills without learning a programming language. However, the students are under limitations such as subscription and other financial requirements, and shortage of deep analysis tools. The second option, on the other hand, opens new horizons for students to carrying out more tailored and in-depth analysis of language through the use of language programming. This option is time-demanding and requires students to be ready to learn at least one programming language. This makes it difficult to learn CL in only one course. The course "Introduction to Computational Linguistics" is going to be the first course of a multi-course module in CL as result of interdisciplinary synergy between linguists and computer scientists. (Chen, X., et al., 2017)

Computational Linguistics and General Linguistics

Linguistics did not attain a complete perfection because of the complex nature of human languages and it is difficult to capture the entire linguistics knowledge with hundred percent accuracy in processing, if possible. The development of the tools for the automatic processing of the natural languages has vital significance in the overall development of any country. And it is equally inevitable to compete with each other globally (Sakthi Vel, 2017). Computational linguistics is closely connected with applied linguistics and linguistics in general. Computational linguistics might be considered as a synonym of automatic processing of natural languages, since the main task of computational linguistics is to construct computer programs to process speech and texts in natural languages (Sakthi Vel, 2017). This, therefore, justifies the recent displacement of CL from Computer Science programs to human language and linguistics programs in some universities such as: University of British Columbia, Georgetown College, Yale University, University of British Columbia, University of Florida, Tübingen University, University of Washington

and other well-known universities. It is due to their interest with the linguistic main component of the module. Other Universities focus on the engineering aspect of the module so that they keep it within Computer Science Programs that I would rather name it '*Linguistic Computation*' than the existing '*Computational Linguistics*'.

The Study Findings

The following sections are dedicated to achieving the objectives of this study. That is, the forthcoming two sections give a clear account to the feasibility of teaching CL in the programs of English department through a review to similar programs in well-known international academic programs either at the BA or the MA levels as well a field survey to the academicians from several academies and universities in Yemen, and students from University of Science and Technology, Yemen. The third section presents a detailed description to the course items that are proposed to be taught withing the two-track description, namely the ready-made software and websites, and the programing track.

Mapping Review of Similar Academic Programs

As mentioned earlier, the course "*Introduction to Computational Linguistics*" is mainly within the course programs of Faculty of Computer Science to deal with its engineering nature. Due to the focus on its salient linguistic element, some universities include this course within the programs of Language Sciences. Figure (1) demonstrates 36 academic programs in 18 countries.

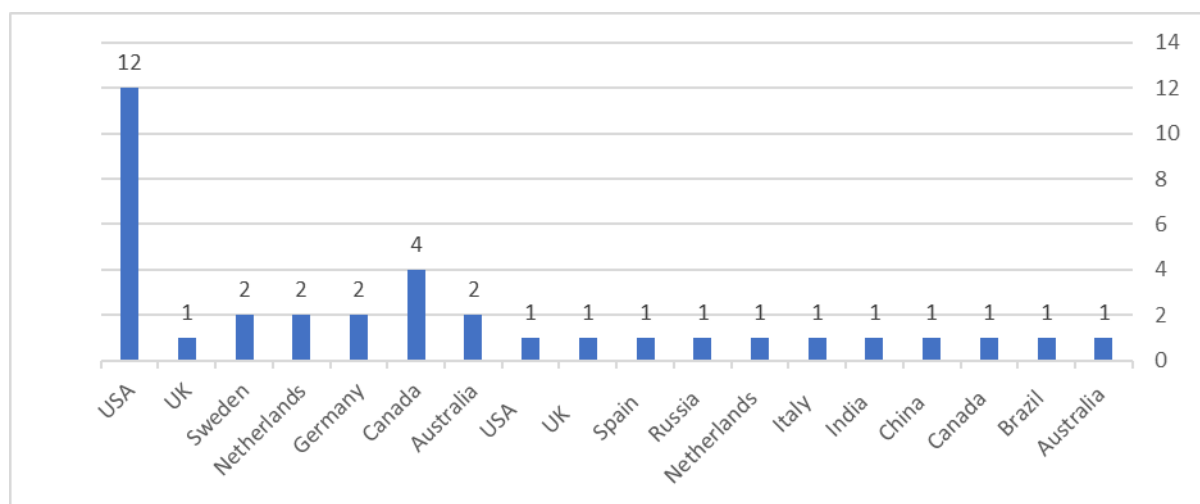


Figure 1: An overview for some countries that have courses named "*(Introduction to Computational Linguistics or Natural Language Processing)*" within the programs of English Department or a relevant Dept.

These programs include the course "*Introduction to CL*" within either the undergraduate level or at the graduate program. The followings are detailed examples:

- Linguistics Department at University of Arizona has introduced an MA degree in Computational Linguistics within its programs. Students must complete the General Education and other degree requirements applicable to the College of Social and Behavioral Sciences and an additional 36 or 39 units of major coursework depending on the year admitted (admit terms Fall 2022 and later require 39 units). (<https://linguistics.arizona.edu>, 2022)
- The School of Linguistics & Applied Language Studies, Wellington Faculty of Humanities and Social Sciences at Victoria University of Wellington, New Zealand has included this course within its program named “*Special Topic: Introduction to Computational Linguistics*”, with code (LING 226) to be taught in 200 hours (15 hours for initials, 20 hours for recorded lectures, and the remaining hours are for reading and assessment) (Victoria University of Wellington, 2022).
- Department of Linguistics at University of Pittsburgh, USA, has introduced a course named “*Computational Linguistics*” code (LING 1330). This course include teaching basics of Python with it NLTK library in addition to introducing real-world applications of computational linguistics: spell checking, part-of-speech tagging, parsing, document classification, and more (<https://www.linguistics.pitt.edu>, 2022).
- Linguistics Department, Faculty of Arts and Sciences at Northwestern University introduces a course named “*Introduction to Computational Linguistics*”, code (LING 334-0) for which basic programming experience at least LING 300-0 (Intro to Text Processing and Programming for Linguists) or COMP_SCI 110-0 (Intro to Computer Programming) are required (<https://catalogs.northwestern.edu/>, 2022).
- At regional level, the program of Applied Linguistics at American University of Cairo, Egypt introduced a course named “*Introduction to Computational Linguistics*”, code (LING 000/5124). This course introduces students to the main concepts of the field and its real-world applications, including, but not limited to, machine translation and information retrieval. Furthermore, it gives students hands-on experience with using and developing computational linguistics tools such as part-of-speech taggers, morphological analyzers, syntactic parsers, and semantic interpreters. To use and develop such tools, students will learn about regular expressions, programming for text analysis, and machine learning (<https://catalog.aucegypt.edu>, 2022).

CL in the Perspective of Yemeni Stakeholder

The study, in this regard, depended on a survey distributed to a sample of the translation stake-holders (43 respondents) such as:

- Sixteen academicians who are from several departments and programs (translation, arts, education and applied linguistics), and have different academic degrees (Demonstrators, MA, and PhD). It is worth noting that 4 academicians were from faculty of Computer science,
- Eighteen undergraduates and post graduates from several English programs, and
- Nine translators who graduated from English department except two of them who graduated from department of Administration and faculty of law.

Table (1) gives a clear account to the sample that was categorized differently according to age, profession, specialization, academic degree and years of experience. The section of data analysis and explanation explains these potential variables and give a clear account to the demographic distribution of the sample categories. The value of T of these variables is between (2.167) and (2.920) which was statistically significant at level of Significance less or equal to (0.05).

Table 1: demographic distribution of sample according to age, specialization, academic degree and experience

Categories	Sub-categories	Freq.	St. deviation	Agreement means	Agreement percentage	Sig.
Age	23-30 yrs	12	0.81	3.3	66%	*0.004
	31-50 yrs	13	0.72	3.7	74%	
	Older than 50 yrs	6	0.58	3.4	68%	
profession	Academicians	16	0.81	3.5	70%	*0.007
	Translators	9	0.66	3.3	66%	
	Students	18	0.72	3.6	72%	
specialization	translation	21	0.79	4	80%	*0.032
	Arts	5	0.81	3.3	66%	
	education	11	0.66	3.7	73%	
	computer science	4	0.72	2.6	52%	
	others	2	0.79	3.7	74%	
Academic degree	PhD	12	0.81	3.4	68%	*0.005
	MA	10	0.66	3.3	66%	
	BA	21	0.72	3.7	74%	
Experience	More than 10 yrs	15	0.79	3.5	69%	*0.025
	6-10 yrs	12	0.81	3.4	68%	
	1-5 yrs	16	0.66	3.6	71%	
		43		3.5	69%	*0.015

(*) Statistically significant at the level of significance ($p \leq 0.05$)

Data analysis and explanation

The table above gives a clear account to the sample and its categories (profession, specialization, academic degree and years of experience)

1. Out of forty-three, forty-two respondents (97.7%) agreed with including CL within all programs of humanities towards achieving digital humanities in general and withing programs of English Department in particular, while only one respondent (2.3%) declared that CL should not be taught withing these programs or any level whatsoever.
2. Forty-two respondents (97.7%) agreed that students should be equipped with enough technical skills and good mastery of using computers before they start with CL. Moreover, they agreed on the necessity to equip the English programs with qualified cadre who can instruct and give professional training, and install computer labs for this course. One respondent (2.3%) chose “Undecided”.
3. Out of forty-two respondents who agree to include CL withing English programs, thirty-two respondents (74.4%) agreed with including CL within BA programs of English department that includes: Translation, Applied Linguistics, Literature, Arts and Education, twenty-three (53.4%) of which agreed to include CL as an elective course, and nine respondents (20.9%) recommended to include CL with BA English programs as and a mandatory course.
4. Out of forty-two respondents, only twenty respondents (46.4%) recommended including CL within MA programs of English Department, four respondent (9.2%) of whom agree on teaching CL as an elective course and sixteen (37.2%), on the other hand, agreed that CL should be taught as a mandatory course withing MA English Programs. It’s worth noting that, after reviewing this result, the researcher tries to investigate the reasons of this disagreement. It was found that three professors whose age is above 50 and did not teach any technology-based courses at all, and 7 respondents were graduates and post-graduates who never have any interest in technology in general and had only 1-5 years of experience in translation.
5. Out of forty-two respondents who agreed to include CL within English programs, thirty-seven respondents (86%) recommended to include CL within only some programs most of which are Translation and Applied Linguistics, and they did not encourage teaching CL withing programs such as Education and Arts.
6. Out of forty-three, thirty-eight (88.3%) agreed that CL can be an independent program within MA level, whereas four respondents (9.3%) chose “Undecided”.

7. Out of forty-three, eighteen respondents (41%) recommended ready-made software to be included within the syllabus of CL, whereas twenty-five respondents (59%) ticked on the “undecided” square.
8. Out of forty-three, fifteen respondent (34.8%) recommended teaching a programming language, such as Python, within the syllabus of CL, whereas twenty-eight respondents (65%) ticked on the “undecided” square.
9. Eighteen respondent (41.8%) agreed to include language processing techniques such as: tokenizing, tagging, parsing and sentiment analysis withing the syllabus of CL whereas twenty-five respondents (59%) ticked on the “undecided” square.
10. It is clear that undecided items, the last items, may be referred to the reason of that respondents may feel that they are foreign to course syllabus of CL, i.e., those who decided on the choice of type of syllabus items expressed their ignorance of this new syllabus to them.
11. Moreover, it’s noticeable that those who chose “undecided” or “disagree” are older than 50 years and those who have poor experience in translation (1-5 years).
12. Besides, instructors and students who majored in translation have unanimously agreed to include CL in English programs either as a course (elective or mandatory), and/or as an independent program at the MA level.

A Course content proposal

As mentioned earlier, there is a two-track proposal to teaching CL in the academic programs in the English Department. These two tracks are elicited from several programs mentioned earlier in some countries like USA, Canada and Britain.

Firstly: Student may take ready-made software, applications and website services to carry out the intended natural language processing (NLP). This product-oriented approach is easier and may equip students with the required NLP skills, they are under the limitations of the service providers and programmers so that they may sometimes face challenges to toiler them to the scope of their assignment such as the limitation of libraries the software may depend on, this can be taken as a course at the undergraduate level such as the American University of Cairo who provide a course name *Introduction to Computational Linguistics*. The following are the major applications that CL can deal with and from which a syllabus may be elicited.

Applications of Computational Linguistics

The proliferation of AI tools such as LLMs that can carry out many NLP processes is a major shift towards automatization of NLP and presented users, who include students, learning new techniques to deal with AI tools such as prompt engineering. According to

Hausser (2014) and Sakthi (2017), the applications of Computational Linguistics are twofold:

I. Language Analysis:

1. *Tokenization* (morphology): Given a character sequence and a defined document unit, tokenization is the task of chopping it up into pieces, called *tokens* (Standford, 2022). Tokenization and sentence splitting can be described as 'low-level' text segmentation which is performed at the initial stages of text processing (Voutilanen, 2005).
2. *Tagging* (syntax): A more informal term for the act of applying additional levels of annotation to corpus data (Baker, Hardie, & McEnery, 2006). The most salient type is Part of Speech Tagging (POS). Tagging means automatic assignment of descriptors, or tags, to input tokens (Voutilanen, 2005).
3. *Part-of-Speech Tagging (POS)*: A type of annotation or tagging whereby grammatical categories are assigned to words (or in some cases morphemes or phrases), usually via an automatic tagger although human post-editing may take place as a final stage
4. *Parsing* (syntax): When a text is parsed, tags are added to it in order to indicate its syntactic structure (Baker, Hardie, & McEnery, 2006).
5. *Corpus Linguistics*: dealing with corpora which include huge amounts of language tokens demand computational and statistical skills to analyze them through the use of the ready-made applications that may not achieve all the goals the analyst, or through use of programming language (like python) to make a tailored code to achieve his own objectives.
6. *Sentiment analysis* (semantics): human-annotated and automated assessment of the attitudes, emotions and opinions of users; includes the identification and study of subjective information (user surveys, customer feedback forms, news media, social media posts) to extract particular words or phrases in order to understand user tone (positive, negative, neutral) and user sentiment (satisfaction, anger, sarcasm) (Lebert, 2021)

II. Language Generation:

Sakthi, (2017), mentioned the following areas of language generation with some elaborated explanation.

1. *Indexing and retrieval in textual databases*: Text indexing is a preprocessing step for text retrieval. During the text indexing process, texts are collected, parsed and stored to facilitate fast and accurate text retrieval (Ling Liu, 2009). The World Wide

Web (WWW) may also be viewed as a large, unstructured textual database, which demonstrates daily to a growing number of users the difficulties of successfully finding the information desired.

2. *Machine translation*: especially in the European Union, currently with 24 official languages, the potential utility of automatic or even semi-automatic translation systems is tremendous.
3. *Automatic text production*: large and multinational companies which continually bring out new products such as motors, CD players, farming equipment, etc., must constantly modify the associated product descriptions and maintenance manuals. A similar situation holds for lawyers, tax accountants, personnel officers, etc., who must deal with large amounts of correspondence in which most of the letters differ only in a few, well-defined places. Here techniques of automatic text production can help, ranging from simple templates to highly flexible and interactive systems using sophisticated linguistic knowledge.
4. *Automatic text checking*: applications in this area range from simple spelling checkers (based on word form lists) via word form recognition (based on a morphological parser) to grammar or syntax checkers based on syntactic parsers which can find errors in word order, agreement, and style checkers that refer certain logarithms of writing styles and discourse. etc.
5. *Automatic content analysis*: The letter-based information on this planet has been said to double every ten years. Even in specialized fields such as natural science, law, and economics, the constant stream of relevant new literature is so large that researchers and professionals do not nearly have enough time to read it all. A reliable automatic content analysis in the form of brief summaries would be very useful. Automatic content analysis is also a precondition for concept-based indexing, needed for accurate retrieval from textual databases, as well as for adequate machine translation.
6. *Automatic tutoring*: there are numerous areas of teaching in which much time is spent on drill exercises such as the more or less mechanical practicing of regular and irregular paradigms in foreign languages. These may be done just as well on the computer, providing the students with more fun (if they are presented as a game, for example) and the teacher with additional time for other, more sophisticated activities such as conversation. Furthermore, these systems may produce automatic protocols detailing the most frequent errors and the amount of time needed for various phases of the exercise. This constitutes a valuable material for improving the automatic tutoring system. It has led to a new field of research in which the 'electronic text book' of old is replaced by new teaching programs utilizing the

special possibilities of the electronic medium to facilitate learning in ways never explored before.

7. *Automatic dialog and information systems*: these applications range from automatic information services for train schedules via queries and storage in medical databases to automatic tax consulting.

Among the main practical tasks which are being solved with the help of the means of the computational linguistics are machine translation, systems for automatic question answering, text retrieval on some subject, text summarization, error correction, analysis of texts or spoken language for some topic, sentiment or other psychological aspects, dialogue agents for accomplishing particular tasks (e.g. purchases, trip planning or medical advising), systems for better language acquisition and gaining knowledge from text (Johnson, 2011).

Secondly: Students may learn programming so that he may carry out NLP tailored assignment and take part in improving the technical aspects of NLP software from a linguist point of view. This process-oriented track is more challenging and time-demanding though. Some universities teach this course as an introductory course such as University of Pittsburgh. Since it requires more than one course, other universities, on the other hand, opened an independent multi-course program named Computational Linguistics at the MA program such as University of Arizona. The following course plan is a proposal of a course named “*Introduction to Computational Linguistics*” that depends on Hammond’s textbook “*Python for Linguists*”. There is no need to say that prescribers may modify or add the suggested course plan to be aligned with the concerned program ILOs providing that these course outcomes are directed to achieving course objectives in the course description document according to their needs, culture and academic orientations.

No.	Topic	Sub-topics	Th.H	Pr. H.
1	Definitions and importance	1.1 importance of CL 1.2 definitions: Digital humanities – computational linguistics – Artificial intelligence AI – Natural language processing – programing – python – tokenization – tagging – parsing 1.3 Installing and Using Python -Interactive Environment 1.4 IDE -Basic Interactions -Edit and Run	2	4
2	Data Types and Variables	2.1 Assignment 2.2 Variable Names 2.3 Basic Data Types 2.3.1 Numbers 2.3.2 Booleans 2.3.3 Strings 2.3.4 Lists 2.3.5 Tuples 2.3.6 Dictionaries	2	4

		2.4 Mutability		
3	Control Structures	3.1 Grouping and Indentation 3.2 if 3.3 Digression on Printing 3.4 for 3.5 while 3.6 break and continue 3.7 Making Nonsense Items	2	4
4	Input–Output	4.1 Command-Line Input 4.2 Keyboard Input 4.3 File Input–Output 4.4 Alice in Wonderland	2	4
5	Subroutines and Modules	5.1 Simple Functions 5.2 Functions That Return Values 5.3 Functions That Take Arguments 5.4 Recursive and Lambda Functions 5.5 Modules 5.6 Writing Your Own Modules 5.7 Docstrings and Modules 5.8 Analysis of Sentences	2	4
6	Regular Expressions	6.1 Matching 6.2 Patterns 6.3 Backreferences 6.4 Initial Consonant Clusters	2	4
7	Mid-term Test		1	2
8	Text Manipulation	7.1 String Manipulation Is Costly 7.2 Manipulating Text 7.3 Morphology	2	4
9	Internet Data	8.1 Retrieving Webpages 8.2 HTML 8.3 Parsing HTML 8.4 Parallelism 8.5 Unicode and Text Encoding 8.6 Bytes and Strings 8.7 What Is the Encoding? 8.8 A Webcrawler	2	4
10	Objects	9.1 General Logic 9.2 Classes and Instances 9.3 Inheritance 9.4 Syllabification	2	4
11	GUIs	10.1 The General Logic 10.2 Some Simple Examples 10.3 Widget Options 10.4 Packing Options 10.5 More Widgets 10.6 Stemming with a GUI	2	4
12	Functional Programming	11.1 Functional Programming Generally 11.2 Variables, State, and Mutability	2	4

			11.3 Functions as First-Class Objects		
			11.4 Overt Recursion		
			11.5 Comprehensions		
			11.6 Vectorized Computation		
			11.7 Iterables, Iterators, and Generators		
			11.8 Parallel Programming		
			11.9 Making Nonsense Items Again		
13	Introduction to NLP	to	12.1 Installing NLTK	2	4
			12.2 Corpora		
			12.3 Tokenizing		
			12.4 Stop Words		
14	Introduction to NLP (2)	to	12.5 Tagging	2	4
			12.6 parsing		
			12.7 Sentiment Analysis		
Final exam				1	2

Discussion

- Universities and other academic institutions are strongly recommended to include CL in the academic programs of English Departments at both undergraduate and postgraduate levels. This course can include ready-made applications, software and websites services that introduce NLP services, or a programming course that can be introductory to a multi-course module at the MA level.
- When setting their pedagogical objectives, the academic programs should include ICT competence as a professional skill for their English academic programs at the undergraduate and postgraduate levels. The students, accordingly, should be equipped with a network that gives them access to high profile e-libraries and think tanks. They also should qualify instructors to text processing tools and applications, and language programming.
- Due to the importance of the practical and professional aspects of the course, four contact hours were allocated for the practical in comparison to only two theoretical hours. However, the program may redistribute the hours according to their potentials and available capacities.
- Students are strongly recommended to be able to have a good mastery of computer application such as basic knowledge of PC components, MS-Office package and internet surfing skills. This is reflected in the syllabus through the pre-requisite line.
- Since students of language-majored programs are not specialized in computer-related sciences, they may take this module as an elective module that gives students who are not interested in this area to take another module such as interpreting or Audiovisual Translation (AVT) for students of translation, and

Comparative Rhetoric for students of Applied Linguistics. Students may learn prompt engineering to communicate with LLMs and other AI tools to avoid hallucinations in their responses.

- Before starting every class, students need to get more focused on Computational Thinking (CT) so that they can understand visualize and analyze programming codes better.
- It might first be helpful for us to just skim through some very fundamental concepts and concepts in Python (Panggabean, 2015). Since it's an introductory course, the fourteen lectures (may be divided into forty-two sessions or twenty-eight-hour sessions at least) to cover all the necessary principles of python programming so that it can be a solid foundation based on which Natural Language Processing NLP could be prescribed (for linguists) in the next course in the module which might be prescribed by any specialized researcher.

Conclusion

At last, definitely not least, being prescribed for language programs, Introduction to Computational Linguistics would be a laying foundation for digital humanities in general, and Natural Language Processing and computational Linguistics modules in particular. This course is introductory to linguistics programing, as well as for linguistic data collection and analysis. This paper can be regarded as a threshold to officially computerizing and digitalizing human and language sciences for a considerable number of academies and universities in developing countries in particular. This implies working on proposals for the other courses in the module of *Computational Linguistics* providing that it fits students of the programs of language departments.

Bibliography List

- Baker, P., Hardie, A., & McEnery, T. (2006). *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press Ltd. doi: <https://doi.org/10.1515/9780748626908>
- Britannica. (2022, 2 2). <https://www.britannica.com/event/Stone-Age>. Retrieved from Stone Age: <https://www.britannica.com/event/Stone-Age>
- Chen, X., Wang, L., & Liu, Q. (2017). *Teaching Computational Linguistics: Integrating Linguistic Theories into the Classroom*. *Language Education and Technology*, 22(3), 215-230.
- Fasold, R., & Connor-Linton, J. (2013). *An Introduction to Language and Linguistics* (Vol. 6th). New York: Cambridge University Press. doi: <https://doi.org/10.1017/cbo9781107707511>
- Hammond, M. (2020). *Python for Linguists*. Cambridge: Cambridge University Press. doi: <https://doi.org/10.1017/9781108642408>
- Hausser, R. (2014). *Foundations of Computational Linguistics*. New York: Springer Heidelberg. doi: <https://doi.org/10.1007/978-3-642-41431-2>

- <https://catalog.aucegypt.edu>. (2022, March 9). Retrieved from American University of Cairo: Courses: https://catalog.aucegypt.edu/content.php?filter%5B27%5D=-1&filter%5B29%5D=5124&filter%5Bcourse_type%5D=-1&filter%5Bkeyword%5D=&filter%5B32%5D=1&filter%5Bcpag%5D=1&cur_cat_oid=20&expand=&navoid=841&search_database=Filter#acalog_template_course_filter
- <https://catalogs.northwestern.edu/>. (2022, March 9). Retrieved from Northwestern: ACADEMIC CATALOG: <https://catalogs.northwestern.edu/undergraduate/courses-az/ling/>
- <https://linguistics.arizona.edu>. (2022, March 9). Retrieved from Linguistics Undergraduate Major Requirements: <https://linguistics.arizona.edu/linguistics-undergraduate-major-requirements>
- <https://www.linguistics.pitt.edu>. (2022, March 9). Retrieved from Department of Linguistics: <https://www.linguistics.pitt.edu/undergraduate/courses>
- Ivashkevych, L. (2019). TEACHING PROGRAMMING WITH PYTHON FOR LINGUISTICS STUDENTS: WHYS AND HOW-TOS. *Advanced Linguistics*, pp. 4-24.
- Jones, C., & Brown, M. (2018). *Enhancing Language Programs: The Role of Computational Linguistics in Linguistic Education*. *Journal of Applied Linguistics*, 45(4), 321-335.
- Johnson, M. (2011). Linguistic Issues in Language Technology. *How relevant is linguistics to computational linguistics*, 6(7), pp. 1-23.
- Kubrický, J., & Částková, P. (2014). Teachers ICT Competence and Their Structure as A Means of Developing Inquiry-Based Education. 5th World Conference on Learning, Teaching and Educational Leadership (pp. 882-885). *Procedia - Social and Behavioral Sciences*.
- Lebert, M. (2021, April 16). *Artificial intelligence (AI) — glossary*. Retrieved from <https://marielebert.wordpress.com/?s=Sentiment+analysis+>
- Levshina, N. (2015). *How to do Linguistics with R: Data exploration and statistical analysis*. Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Ling Liu, M. T. (Ed.). (2009). *SpringerLink*. Retrieved from Encyclopedia of Database Systems: https://link.springer.com/referenceworkentry/10.1007/978-0-387-39940-9_417
- Muthanna, A., & Karaman, A. (2014). Higher education challenges in Yemen: Discourses on English Teacher Education. *International Journal of Educational Development*, 40-47. doi: <http://dx.doi.org/10.1016/j.jedudev.2014.02.002>
- *NATIONAL STRATEGY FOR THE DEVELOPMENT OF HIGHER EDUCATION IN YEMEN*. (2005), Sana'a: Ministry of Higher Education and Scientific Research.
- Panggabean, H. &. (2015). *Journal of Humanities and Social Science (IOSR-JHSS)*. *Computational Linguistics Application Using Python Programming*, 20(7), pp. 18-30.
- Roth, B., & Wiegand, M. (2021, April 21). *Computational Linguistics*. *MIT Press Direct*, pp. 217-220. doi: https://doi.org/10.1162/coli_r_00400
- Sakthi Vel, S. (2017). Applications of Computational Linguistics to Language Studies: An Overview. *International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)*, 4(3), pp. 239-244.
- Smith, A., & Johnson, B. (2016). *The Synergy of Linguistics and Computational Linguistics: A Call for Collaboration*. *Computational Linguistics Journal*, 40(2), 112-128.

- Stanford, U. o. (2022, 2 16). *index*. Retrieved from <https://nlp.stanford.edu/IR-book/html/htmledition/index-1.html>
- *Victoria University of Wellington*. (2022, March 9). Retrieved from Course offering: <https://www.wgtn.ac.nz>
- Voutilanen, A. (2005). *The Oxford Handbook of Computational Linguistics*. In R. Mitkov (Ed.). Oxford.