ASU

GENERAL STUDIES COURSE PROPOSAL COVER FORM

Course information:

Copy and paste __current__ course information from Class Search/Course Catalog.

College/School	College of Liberal Arts and Sciences		Department	**SILC**	

Prefix	**SLV**	Number	**304**	Title	Computational linguistics of Slavic Languages	Units:	3

Is this a cross-listed course?	No	If yes, please identify course(s)

Is this a shared course?	No	If so, list all academic units offering this course

Note- For courses that are crosslisted and/or shared, a letter of support from the chair/director of __each__ department that offers the course is required for __each__ designation requested. By submitting this letter of support, the chair/director agrees to ensure that all faculty teaching the course are aware of the General Studies designation(s) and will teach the course in a manner that meets the criteria for each approved designation.

Is this a permanent numbered course with topics?	Yes

If yes, all topics under this permanent numbered course must be taught in a manner that meets the criteria for the approved designation(s). It is the responsibility of the chair/director to ensure that all faculty teaching the course are aware of the General Studies designation(s) and adhere to the above guidelines.

Chair/Director Initials

_____
(Required)

Course description:

Requested designation: Mathematical Studies–CS

Note- a __separate__ proposal is required for each designation requested

Eligibility:

Permanent numbered courses must have completed the university's review and approval process.
For the rules governing approval of omnibus courses, contact Phyllis.Lucie@asu.edu or Lauren.Leo@asu.edu.

Submission deadlines dates are as follow:

For Fall 2016 Effective Date: October 6, 2015   For Spring 2017 Effective Date: March 15, 2016

Area(s) proposed course will serve:

A single course may be proposed for more than one core or awareness area. A course may satisfy a core area requirement and more than one awareness area requirements concurrently, but may not satisfy requirements in two core areas simultaneously, even if approved for those areas. With departmental consent, an approved General Studies course may be counted toward both the General Studies requirement and the major program of study.

Checklists for general studies designations:

Complete and attach the appropriate checklist

- Literacy and Critical Inquiry core courses (L)
- Mathematics core courses (MA)
- Computer/statistics/quantitative applications core courses (CS)
- Humanities, Arts and Design core courses (HU)
- Social–Behavioral Sciences core courses (SB)
- Natural Sciences core courses (SQ/SG)
- Cultural Diversity in the United States courses (C)
- Global Awareness courses (G)
- Historical Awareness courses (H)

A complete proposal should include:

☒ Signed course proposal cover form
☒ Criteria checklist for General Studies designation(s) being requested
☒ Course catalog description
☒ Sample syllabus for the course
☒ Copy of table of contents from the textbook and list of required readings/books

It is respectfully requested that proposals are submitted electronically with all files compiled into one PDF.

Contact information:

Name	Danko Sipka	E-mail	Danko.Sipka@asu.edu	Phone	480 634 8427

Department Chair/Director approval: *(Required)*

Chair/Director name (Typed):	Robert Joe Cutter	Date:	8/4/15

Chair/Director (Signature):		8/4/15

Rev. 4/2015

M SLV 304 Computational Linguistics of Slavic Languages. (3)
spring
Information technology and Slavic languages, including Web design, digitalized resources, information retrieval, math/statistical analysis, and PERL. Lecture, lab.

MATHEMATICAL STUDIES [CS]

Rationale and Objectives

The **Mathematical Studies** requirement is intended to ensure that students have skill in basic mathematics, can use mathematical analysis in their chosen fields, and can understand how computers can make mathematical analysis more powerful and efficient. The **Mathematical Studies** requirement is completed by satisfying both the **Mathematics [MA]** requirement and the **Computer/Statistics/Quantitative Applications [CS]** requirement explained below.

The **Mathematics [MA]** requirement, which ensures the acquisition of essential skill in basic mathematics, requires the student to complete a course in College Mathematics, College Algebra, or Pre-calculus; or demonstrate a higher level of skill by completing a mathematics course for which a course in the above three categories is a prerequisite.

The **Computer/Statistics/Quantitative Applications [CS]** requirement, which ensures skill in real world problem solving and analysis, requires the student to complete a course that uses some combination of computers, statistics, and/or mathematics.* Computer usage is encouraged but not required in statistics and quantitative applications courses. At a minimum, such courses should include multiple demonstrations of how computers can be used to perform the analyses more efficiently.

*CS does *not* stand for computer science in this context; the "S" stands for statistics. Courses in computer science must meet the criteria stated for CS courses.

Revised April 2014

Proposer: Please complete the following section and attach appropriate documentation.

		ASU--[CS] CRITERIA	
		A COMPUTER/STATISTICS/QUANTITATIVE APPLICATIONS [CS] COURSE MUST SATISFY ONE OF THE FOLLOWING CRITERIA: 1, 2, OR 3	
YES	**NO**		**Identify Documentation Submitted**
		1. **Computer applications***: courses must satisfy both **a** and **b**:	
☒	☐	a. Course involves the use of computer programming languages or software programs for quantitative analysis, algorithmic design, modeling, simulation, animation, or statistics.	SLV304Criteria
		b. Course requires students to analyze and implement procedures that are applicable to at least one of the following problem domains **(check those applicable)**:	SLV304Criteria
☒	☐	i. Spreadsheet analysis, systems analysis and design, and decision support systems.	SLV304Criteria
☒	☐	ii. Graphic/artistic design using computers.	SLV304Criteria
☒	☐	iii. Music design using computer software.	SLV304Criteria
☐	☒	iv. Modeling, making extensive use of computer simulation.	
☒	☐	v. Statistics studies stressing the use of computer software.	SLV304Criteria
☐	☐	vi. Algorithmic design and computational thinking.	

*The **computer applications** requirement **cannot** be satisfied by a course, the content of which is restricted primarily to word processing or report preparation skills, the study of the social impact of computers, or methodologies to select software packages for specific applications. Courses that emphasize the use of a computer software package are acceptable only if students are required to understand, at an appropriate level, the theoretical principles embodied in the operation of the software and are required to construct, test, and implement procedures that use the software to accomplish tasks in the applicable problem domains. Courses that involve the learning of a computer programming language are acceptable only if they also include a substantial introduction to applications to one of the listed problem domains.

YES	NO		Identify Documentation Submitted
		2. Statistical applications: courses must satisfy **a, b,** and **c.**	
☐	☐	a. Course has a minimum mathematical prerequisite of College Mathematics, College Algebra, or Pre-calculus, or a course already approved as satisfying the MA requirement.	
		b. The course must be focused principally on developing knowledge in statistical inference and include coverage of all of the following:	
☐	☐	i. Design of a statistical study.	
☐	☐	ii. Summarization and interpretation of data.	
☐	☐	iii. Methods of sampling.	
☐	☐	iv. Standard probability models.	
☐	☐	v. Statistical estimation	
☐	☐	vi. Hypothesis testing.	
☐	☐	vii. Regression or correlation analysis.	
☐	☐	c. The course must include multiple demonstrations of how computers can be used to perform statistical analysis more efficiently, if use of computers to carry out the analysis is not required.	

YES	NO		Identify Documentation Submitted
		3. Quantitative applications: courses must satisfy **a, b,** and **c:**.	
☐	☐	a. Course has a minimum mathematical prerequisite of College Mathematics, College Algebra, or Pre-calculus, or a course already approved as satisfying the MA requirement.	
		b. The course must be focused principally on the use of mathematical models in quantitative analysis and decision making. Examples of such models are:	
☐	☐	i. Linear programming.	
☐	☐	ii. Goal programming.	
☐	☐	iii. Integer programming.	
☐	☐	iv. Inventory models.	
☐	☐	v. Decision theory.	
☐	☐	vi. Simulation and Monte Carlo methods.	
☐	☐	vii. Other (explanation must be attached).	
☐	☐	c. The course must include multiple demonstrations of how computers can be used to perform the above applications more efficiently, if use of computers is not required by students.	

Course Prefix	Number	Title	General Studies Designation
SLV	304	Computational Linguistics of Slavic Languages	

Explain in detail which student activities correspond to the specific designation criteria.
Please use the following organizer to explain how the criteria are being met.

Criteria (from checksheet)	How course meets spirit (contextualize specific examples in next column)	Please provide detailed evidence of how course meets criteria (i.e., where in syllabus)
1	See SLV304Criteria	SeeSLV304Criteria

SLV 304 Computational Linguistics of Slavic Languages
Specific CS Criteria Met by the Course

Introduction

This course has been designed for Slavic, and possibly other foreign language majors, seeking to meet the CS general studies requirement while at the same time pursuing research interests in their own field of concentration. Historically, students from the Slavic Section, in command of basic IT skills, have expressed vivid interest in fulfilling the CS general studies requirement by engaging in activities linking IT with their own field of concentration. These legitimate interests are further justified by real-world developments in recent years, where computational linguistics has become one of the most prominent areas of computer science bringing together linguists and computer scientists. This course is being offered as RUS 304 in the spring semester of 2003. Both the course design and the current enrolment of fourteen students offer proof that adding the CS designation **will not change the general pattern of foreign language majors taking CS courses at the School of Computing but rather open another possibility for a relatively small group of students with specific interests**.

Description by Each Criterion

This course meets the set of criteria labeled 1 (Computer applications). Although some criteria from sets 2 and 3 are met in this course, we will provide comments about set 1 only. Each criterion will be addressed in turn.

a. Use of Programming Languages
Four weeks of this course are devoted to programming in Perl. Topics include: branching, looping, arrays, input/output handling (including file input/output), regular expressions, using PERL in cgi scripts, etc.
See: Attached syllabus, Course Web page http://www.asusilc.net/asucomp
(Programming in Perl link)

b. Other

i. Spreadsheet analysis, etc.
Two weeks of this course are devoted to working with databases. After a brief introduction to basic spreadsheet analyses in Excel, students are required to create an electronic dictionary in the form of an Access database and perform sorting, consistency analyses, filtering, importing/exporting, etc.
See: Attached syllabus, Course Web page http://www.asusilc.net/asucomp
(Computational Lexicography link)

ii. Graphic design, etc.
The first four weeks of this course cover Web page design. This involves learning HTML codes, including forms, and JavaScript, as well as general visual design. Students are offered not only an introduction into the general principles of Web page

visual design, but receive also hands-on knowledge from working with different programs ranging from Microsoft Paint to Micromedia Fireworks, and Dreamweaver.
See: Attached syllabus, Course Web page http://www.asusilc.net/asucomp (Web Pages and Code Pages and Interactive On-line Exercises links)

iii. Music design, etc.
Another part of the first four weeks of the course is the incorporation of musical materials into static and interactive Web pages. In addition, students learn how to record and modify sound with both Microsoft Sound Recorder and GoldWave, a more elaborate music design shareware program.
See: Attached syllabus, Course Web page http://www.asusilc.net/asucomp (Web Pages and Code Pages and Interactive On-line Exercises links)

v. Statistics
The course features a section devoted to statistical analysis for linguists. Students are familiarized with basic descriptive statistics as applied to the study of language and also with correlation coefficient as an example of inferential statistical analysis. They are furthermore required to use SPSS – to code and import their data, perform the analysis and tabulate/graph the results. In addition, the two weeks of corpus linguistics comprise frequency analysis, z scores, etc.
See: Attached syllabus, Course Web page http://www.asusilc.net/asucomp (Statistics Analysis and Textual Corpora links)

Please note: Criterion iv. is not mentioned here as the course does not cover classical modeling. It is true, however, that in the Digitizing Printed Resources section of the course, students use one of the first several pages as a sample to train Optical Character Recognition (OCR) software and SGML/XML tag the text which is then applied to the entire text

SLV 304 Computational Linguistics of Slavic Languages:

3 credit hours

Instructor Danko Sipka, Ph.D.; Associate Professor Research; e-mail: danko.sipka@asu.edu; Web: http://www.public.asu.edu/~dsipka

Course Web Page http://www.asu.edu/clas/reesc/asucomp

Time Spring semester 2014, TuTH, 3:00-4:10 p.m; Office hours TuTh 1:30-3:00

Prerequisites The only prerequisite is familiarity with passive utilization of Microsoft Word and Microsoft Explorer. The course is not limited to Slavists. Slavic languages are going to be used only as an example. Participants from other language and literature fields are welcome.

Objectives This information technology certificate course has the two following principal objectives: a) to develop basic skills required of computational linguists thus enabling participants to be competitive in language industries, and b) to create the foundations for possible further development in more advanced fields in language industries.

Description This course focuses on the following fields of computational linguistics: a) computer-assisted language learning, CALL, (Web pages, Unicode and other standards, interactive on-line exercises), b) digitizing (scanning, optical character recognition, TEI and LMF formatting), c) textual corpora (concordances, frequency analysis, reversed lexical lists, etc.), d) statistical analysis (descriptive and elementary inferential statistics), e) data and text mining in the Web (search engines, Web resources, Web crawlers), f) computational lexicography (lexical databases, morphological taggers and parsers), g) programming in Perl and Python (branching, looping, arrays, file manipulation, regular expressions).Students are assigned several practical tasks within each field. While emphasizing a hands-on component, this course will also familiarize its participants with the basic concepts in the fields of language industries. Following are details on the course topics and the sequence in which they will be presented.

Overview
Course mechanics; Structural characteristics of Slavic languages and their relevance to computational linguistics; Language industries: possibilities and limitations; Operating systems (Windows, Mac OS, Linux) and non-English writing systems.
HW # 1 (due end of Week 1): *Part 1: Get your asurite id if you do not have one, create a Web space at www.public.asu.edu and e-mail your personal space address to me. Part 2: Set-up your computer in a manner such that you can create a text in a non-Western-European script. E-mail me such text.*

Web pages and code pages
HTML coding; Representing various code pages with special emphasis on Unicode; Formatting; Tables; Links; Multimedia content (images, movies, sound), and other bells and whistles, using ready-made software to design Web pages.

HW # 2 (due end of Week 2): *Create a Web page in your personal space which contains non-English characters, tables, images (e.g., your mug shot), sounds (e.g., you cursing in an exotic language), and links. E-mail the link to your page to me*

Interactive on-line exercises
CALL; HTML forms; Java Script; Exercise formats (multiple choice, fill in the blank, etc.); Blackboard & Co.; More bells and whistles with JavaScript.
HW # 3 (due end of Week 4): *Part 1: Create two Web pages each with ten exercise tasks and each with a different exercise format, Part 2: Design a Web page with dynamic multimedia content using tables and OnMouseOver/OnMousOut function. E-mail the links to all these pages to me.*

Digitizing printed resources
Scanning; OCR packages with non-Western character support (FineReader, Recognita, Cuneiform, etc.); SGML/XML formatting; TEI standard.
HW # 4 (due end of Week 6): *Scan one page from a non-Western script dictionary; use an OCR software package to convert it into a text. SGML code the first ten entries using the TEI standard. E-mail both the pure text and the SGML coded text to me.*

Working with textual corpora
Representativeness; Practical applications in language corpora; Concordances; Frequency lists; Reverse lists; Content analysis.
HW # 5 (due end of Week 8): *Create two corpora from two different newspaper sections (e.g., Sports and World) from a newspaper of your choice. Download at least 500K of text from each section. Create a concordance and frequency list for each corpus. Analyze the data and compose a report stating your findings. Attach the top 50 ranks from the frequency list to your report and e-mail it to me.*

Statistical analysis
Data gathering and coding; Working with statistical packages; Descriptive statistics (frequency, mean, standard deviation, etc.); Inferential statistics (correlation coefficient); Graphic presentation of the results.
HW # 6 (due end of Week 9): *Write a research proposal utilizing correlation coefficient as a tool to test a hypothesis related to your field of specialization. E-mail it to me.*

Data mining in the Web
Web search engines; Web resources for Slavic language and area studies; General linguistic resources; Web crawlers.
HW # 7 (due end of Week 10): *Create a Web page devoted to a specific problem with a short description and at least thirty relevant external links. Send the link to your Web page to me.*

Computational lexicography and grammar
Natural language processing (NLP) of Slavic languages; Lexical databases; Sorting; Exporting and importing; Morphological parsers; Morphological generators.
HW # 8 (due end of Week 12): *Create a 50- entry bilingual dictionary (non-Western language X - English) with equivalents, POS tags, usage labels, examples and their translation in the form of a Microsoft Access database. E-mail both the database and the English – non-Western language X index as a text file exported from the database.*

Programming in PERL and Python
Branching; Looping, Arrays, File manipulation; Regular expressions; Perl and CGI Scripts.
HW # 9 (due end of Week 13) *Write a Perl or Python script which converts any input string by replacing all voiceless consonants into voiced ones and vice versa. E-mail the script to me.*
HW # 10 (due end of Week 14) *Write a Perl script which extracts suffixes entered by the user from any given textual corpus. E-mail the script to me.*
Final Project (due end of Week 16): *Export your dictionary from HW # 8 into a text file. Write a HTML form and a Perl code to be used as a CGI script which will query your database according to the headword, the equivalent, POS tag and the usage label. E-mail the dictionary text file, the html form, and the Perl cgi script to me.*

Course Materials
All course materials, including tutorials for further exploration of each of the covered topics, will be available on-line from the course Web page.

Grading Policy
Student performance is evaluated through weekly projects and the final project (see Description above). All projects must be submitted in order to pass the course. Students who submit all projects in a timely manner and in the correct form will receive A. The grade will be lowered one level if there are two late projects or two incomplete projects. Thus, one late/incomplete project is still A, two or three late/incomplete projects is B, four or five late/incomplete projects is C, six or seven late/incomplete projects is D, eight or more late or incomplete projects is E. The final project must be submitted in a timely manner. Incompletion of the final project counts as two incomplete projects.

Schedule
Weeks 1 and 2: Overview, Web pages and code pages; Weeks 3 and 4: Interactive on-line exercises; Weeks 5 and 6: Digitizing printed resources; Weeks 7 and 8: Working with textual corpora; Weeks 9 and 10: Statistical analysis, Data mining in the Web; Weeks 11 and 12: Computational lexicography; Weeks 13-16: Programming in PERL

SLV 304 Textbook Table of Contents

This required documentation is not applicable to SLV 304 as this is a paperless course. All courseware is available in electronic format through the course Web page <http://www.asu.edu/clas/reesc/asucomp>. All student assignments are given and grading is performed through the Web page.

Some examples of the links used throughout the course are as follows:

http://www.w3c.org/MarkUp/Guide - Getting started with HTML
http://www.w3schools.com/default.asp - XML, JavaScript
http://www.tei-c.org – SGML, TEI standard
http://www.linguistics.ucla.edu/people/Kracht/courses/mathling2/statistics.pdf - Statistics
http://www.bu.edu/linguistics/UG/course/lx865/handouts/intro_to_perl_programming.pdf - PERL
http://www.nltk.org/book/ - Python