GENERAL DETAILS OF THE SUBJECT COURSE

- **Name**: **Computational Syntax**
- **Mode**: compulsory
- **Languages**: English

CONTEXT AND KEY QUALIFICATION:

Context:

This is a compulsory .4.5-credit course in the HAP/LAP master. It will be presented during the first semester of the course.

Preconditions:

This course does not have any special precondition. These are the links with other subjects:

- It is recommendable to have a basic knowledge of the basic concepts of Linguistics (those that were taught in secondary school). It is also helpful to know the concepts presented in this master's “Foundations of Linguistics” course
- Programming: it is interesting to understand the basic concepts of programming. Related to this, the course on “Introduction to programming” or a similar one is recommendable
- It is helpful to have a basic understanding of regular expressions

These are the **main objectives** of this course:

- Presentation of the main approaches for the computational treatment of syntax. Among others, context-free grammars, finite-state syntax, and statistical models. Several formalisms will be presented. We will also present a main overview of the main algorithms for Tagging, Chunking and Parsing.
- Overview of the role that computational morphology and syntax play on the applications that make use of language technology.

TEACHING STAFF

GOJENOLA GALLETEBEITIA, KOLDOBIKA
PROFESOR TITULAR DE UNIVERSIDAD
PhD
Computer Languages and Systems
Euskal Herriko Unibertsitatea
koldo.gojenola@ehu.eus

BARNES, JEREMY C.
PROFESOR
Phd
Computer Languages and Systems
Euskal Herriko Unibertsitatea
jeremy.barnes@ehu.eus

URIZAR ENBEITA, RUBEN
PROFESOR TITULAR DE UNIVERSIDAD
PhD
Teaching of Language and Literature
Euskal Herriko Unibertsitatea
ruben.urizar@ehu.eus

SKILLS (Name/weighting)

15670 - Ability to handle, enrich and use language resources for the processing of human language (20%)

15671 - Understanding of the basic strategies for the analysis of language, and capacity of extending these strategies for their use in applications for language processing (20%)

15672 - Ability to use and adapt the tools (morphological, syntactic and semantic analyzers) available for different languages (20%)

15673 - Ability to design and develop resources, tools and computer applications for language technologies (20%)

15674 - Ability to use and adapt the relevant methods for research on language technologies (20%)

TYPES OF TEACHING (Type/Face-to-face hours/Non face-to-face hours/Total hours)

Type	Face-to-face hours	Non face-to-face hours	Total hours
Lecture	26	19	45
Seminar	4	18	22
Exercises (classroom)	4	11	15
Exercises (computer)	12	19	31

TRAINING ACTIVITIES (Name/Hours/ % of classroom teaching)

Name	Hours	% of classroom teaching
Expositive classes	46	58

Discussion	22	16
Exercises	15	24
Application Workshops	30	38

ASSESSMENT SYSTEMS (Name/Minimum weighting/Maximum weighting)

- Attending lectures and active participation. (5%)
- Each student should present exercises and work on each part (75%)
- At the end of the course (see the last day in the course schedule), there will be a written exam. Students will be allowed to use the material presented during the course (20%)

ASSESSMENT CRITERIA (weight)

1. Lectures
   a) Attending the lectures (1 point)
   b) Taking part actively (2 points)
   c) Sound justifications for comments (2 points)

2. Exercises
   a) Solving all the exercises (1 point)
   b) Giving correct results (2 points)
   c) Sound justifications for results (2 points)

3. Work
   a) Correct description of the performed work (1 point)
   b) Correct use of programs/formalisms for the task at hand (2 points)
   c) The data and results are appropriately described and documented (1 point)
   d) The documentation is adapted to the communication context (1 point)

4. Exam
   a) Correct answers (2 points)
   b) Grounded explanations (2 points)
   c) Correct writing and presentation (1 point)

ASSESSMENT SIGNATURE(S)

1 (a) – Did not attend any lecture (0 points)
– Attended most of the lectures (0,5 points)
– Attended all of the lectures (1 points)
(b) – Did not take part at classes (0 points)
– Did take part at classes only a few times (1 points)
– Take active part in class (2 points)
(c) – Did not make any comment on the lectures (0 points)
– The comments were not clearly presented (1 points)
– The comments were clearly presented, also giving new ideas(2 points)

2 (a) – Did not do any exercise (0 points)
– Did most exercises (0.75 points)
– Did all exercises (1 points)
(b) – Only a few solutions were correct (0.5 points)
– Most of the solutions were correct (1 points)
– All of the solutions were correct (2 points)
(c) – The explanations did not match with the results (0 points)

- Several results were not clearly presented (1 points)
- The results were clearly presented and new proposals were made(2 points)

3 (a)
- The work does not give a clear indication of the followed procedures (0 points)
- There are several details missing about the followed procedures (0.5 points)
- The work gives a clear indication of the followed procedures (1 points)

(b)
- The work did not make a correct use of the tools (0.5 points)
- The tools are used correctly, but the work does not evaluate the quality of the data (1 points)
- The tools are used correctly, and the work does evaluate the quality of the data (2 points)

(c)
- Data and results are not documented clearly (0.25 points)
- Most data and results are documented clearly, but they are not compared to similar datasets (0.5 points)
- All data and results are documented clearly, they are compared to similar datasets and presented clearly (1 points)

(d)
- The documentation is not suited to the communication setting. The text does not follow a clear presentation and some information is missing (0.25 points)
- The documentation is correct with respect to gender, style and register. The arrangement of the ideas can be improved (0.5 points)
- The text is arranged clearly, using the correct resources and terminology (1 points)

4 (a)
- Only a few answers are correct (0.5 points)
- Most of the answers are correct (1 points)
- All answers are correct (2 points)

(b)
- The explanations are confusing (0.5 points)
- The explanations are superficial (1 points)
- The explanations are correctly motivated (2 points)

(c)
- There are misspellings and the used terminology is not appropriate (0.25 points)
- It is correctly written, but the organization of ideas is not good (0.5 points)
- The text is correct both formally and regarding its content (1 points)

SYLLABUS (theoretical and practical)

1. Introduction to Computational Syntax
2. Finite-State Syntax
   2.1. N-gram Language Models
   2.2. Assignment of Syntactic Categories (POS tagging)
   Knowledge-based (Constraint Grammar)
   Data-driven (statistical methods)
   2.3. Chunking
3. Multiword expressions (MWE)
4. Context Free Grammars
   Basic model
   Probabilistic Context Free Grammars
   Unification-based Grammars
5. Dependency Syntax
   Rule-based
   Data-driven
6. Syllabus of practical part:

— Exercises related to finite-state automata and context-free grammars
— Examination of examples and resources for the understanding of basic concepts related to computational syntax

- Interpretation of grammar rules
- Development and implementation of grammar rules

BASIC BIBLIOGRAPHY

- Clark A., Fox C., Lappin S. (eds.) (2012). *The Handbook of Computational Linguistics and Natural Language Processing*. Blackwell Handbooks in Linguistics, John Wiley & Sons.
- Eisenstein, J. (2019) *Introduction to Natural Language Processing*, MIT Press Ltd, Cambridge, United States, ISBN10 0262042843, ISBN13 9780262042840
- Hopcroft J., Motwani R., Ullman J. (2001). *Introduction to automata theory, languages and computation*. Pearson-Addison Wesley.
- Jurafsky D., Martin, J. H. (2008) *Speech and Language Processing* (Second Edition), Prentice Hall, Upper Saddle River, N.J.
- Kaplan R. B. (ed) (2010). *The Oxford Handbook of Applied Linguistics*. Second Edition Edited by OUP USA Oxford Handbooks in Linguistics.
- Manning, C. D., Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*, MIT Press Cambridge, Mass.
- Indurkhya, N., Damerau, F. J. (2010). *Handbook of Natural Language Processing*, Second Edition. Chapman & Hall/CRC Machine Learning & Pattern Recognition.
- Roark B. and Sproat R. (2007). Computational Approaches to Morphology and Syntax. Oxford University Press
- Sipser, M. (2012). Introduction to the Theory of Computation. Cengage Learning.
- Smith, N. A. (2011). Linguistic Structure Prediction. Morgan & Claypool Publishers, Synthesis Lectures on Human Language Technologies, 4 (2).

IN-DEPTH BIBLIOGRAPHY

- Aldezabal I., Ansa O., Artola X., Ezeiza A., Gojenola K., Insausti J.M. eta Lersundi M. (1999). *Euskararen Datu-Base Lexikala (EDBL): eskema berriaren proposamena.* Barnetxostena, UPV/EHU/LSI/TR 9-99. Lengoaia eta Sistema Informatikoak Saila, Informatika Fakultatea, Donostia-San Sebastián.
- Aldezabal I., Arriola J.M., Díaz de Ilarraza A. eta Sarasola K. (2005). Hizkuntzalaritza Konputazionala. Udako Euskal Unibertsitatea (UEU), Bilbo.
- Bemova A., Hajic J., Hladka B. eta Panevova J. (1999). Morphological and Syntactic Tagging of the Prague Dependency Treebank. *Journées Atala, Corpus annotés pour la syntaxe*. Paris, France.
- Bick, E. (2000). *The Parsing System 'Palavras': Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*, Aarhus University Press, Aarhus.
- Bird S., Klein E., and Loper E. (2009), *Natural Language Processing with Python --- Analyzing Text with the Natural Language Toolkit*. O'Reilly Media. http://www.nltk.org/book/
- Chomsky, N. (1957). Syntactic structures. The Hague: Mouton.
- Karlsson F., Voutilainen A., Heikkilä J. eta Anttila A. (1995). *Constraint Grammar: A Language-independent System for Parsing Unrestricted Text*. Mouton de Gruyter, Berlin.
- Kiraz G. A. (2001) *Computational Nonlinear Morphology: With Emphasis on Semitic Languages*.
- Socher S., Bauer J., Manning C.D. (2013). Parsing with compositional vector grammars. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*.
- Basic Constraint Grammar Tutorial for CG-3 (Vislcg3). http://beta.visl.sdu.dk/cg3_howto.pdf