Natural Language Processing

10. Coursework

A Galassi F Ruggeri P Torroni

Fall 2022

Notice

Downloading and ~~sharing~~

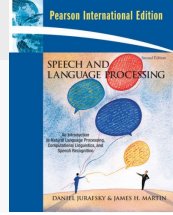
A copy of these slides can be downloaded from virtuale and stored for personal use only. Please do not redistribute.

Table of Contents

Today's programme

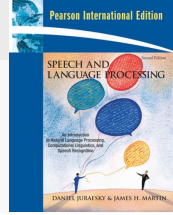
- Brief background for A1: POS tagging (10 minutes)
- A1, Andrea Galassi (15 minutes)
- A2, Federico Ruggeri (15 minutes)
- Short Break
- Project guidelines and evaluation (15 minutes)
- Project report template (5 minutes)
- Standard project (15 minutes)
- Custom projects (5 minutes)

Part-of-Speech Tagging



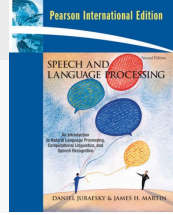
Parts of Speech

- **Parts of speech** notion more than 2,000 years old
- Grammatical sketch of Greek by Dionysus Thrax of Alexandria (*c*. 100 B.C.)
  - noun, verb, pronoun, preposition, adverb, conjunction, participle, article
- **POS/word classes/syntactic categories**
  - useful abstractions
  - reveal a lot about words and their neighbours
- Useful feature in NER, information extraction, co-reference resolution



English Word Classes

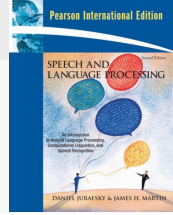
- POS defined based on syntactic and morphological function
  - distributional properties: similar neighbours
  - morphological properties: similar affixes
- Two broad supercategories
  - **closed class** types
    - such as prepositions like *from* and *to*
    - **function words** that tend to be short, frequent, and have structuring uses in grammar
  - **open class** types
    - such as nouns, adjectives and verbs like *blogosphere* and *friendzone*



The Penn Treebank Part-of-Speech Tagset

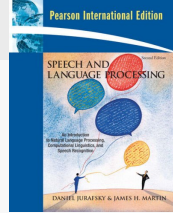
Tag	Description	Example	Tag	Description	Example	Tag	Description	Example
CC	coordinating conjunction	*and, but, or*	PDT	predeterminer	*all, both*	VBP	verb non-3sg present	*eat*
CD	cardinal number	*one, two*	POS	possessive ending	*'s*	VBZ	verb 3sg pres	*eats*
DT	determiner	*a, the*	PRP	personal pronoun	*I, you, he*	WDT	wh-determ.	*which, that*
EX	existential 'there'	*there*	PRP$	possess. pronoun	*your, one's*	WP	wh-pronoun	*what, who*
FW	foreign word	*mea culpa*	RB	adverb	*quickly*	WP$	wh-possess.	*whose*
IN	preposition/ subordin-conj	*of, in, by*	RBR	comparative adverb	*faster*	WRB	wh-adverb	*how, where*
JJ	adjective	*yellow*	RBS	superlatv. adverb	*fastest*	$	dollar sign	*$*
JJR	comparative adj	*bigger*	RP	particle	*up, off*	#	pound sign	*#*
JJS	superlative adj	*wildest*	SYM	symbol	+,%, &	“	left quote	‘ or “
LS	list item marker	*1, 2, One*	TO	“to”	*to*	”	right quote	’ or ”
MD	modal	*can, should*	UH	interjection	*ah, oops*	(	left paren	[, (, {, <
NN	sing or mass noun	*llama*	VB	verb base form	*eat*	)	right paren	],), }, >
NNS	noun, plural	*llamas*	VBD	verb past tense	*ate*	,	comma	,
NNP	proper noun, sing.	*IBM*	VBG	verb gerund	*eating*	.	sent-end punc	. ! ?
NNPS	proper noun, plu.	*Carolinas*	VBN	verb past part.	*eaten*	:	sent-mid punc	: ; ... – -

Figure 8.1 Penn Treebank part-of-speech tags (including punctuation).



Examples

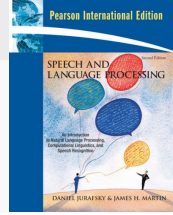
1. The grand jury commented on a number of other topics .
2. There are 70 children there
3. Preliminary findings were reported in today's New England Journal of Medicine.
4. Well, I, I want to go to a restaurant



Examples

1. The/DT grand/JJ jury/NN commented/VBD on/IN a/DT number/NN of/IN other/JJ topics/NNS ./.
2. There/EX are/VBP 70/CD children/NNS there/RB
3. Preliminary/JJ findings/NNS were/VBD reported/VBN in/IN today/NN's/POS New/NNP England/NNP Journal/NNP of/IN Medicine/NNP ./.
4. Well/UH ,/, I/PRP ,/, I/PRP want/VBP to/TO go/VB to/IN a/DT restaurant/NN

For fun: you can play with online POS taggers



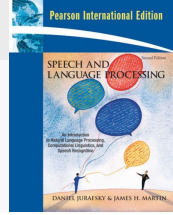
Corpora

Popular corpora (NLTK bundle)

- **Brown**, 1MI words, various genres, US, 1961
- **WSJ**, 1 MI words, 1989
- **Switchboard**, 2 MI words of phone conversations, 1990-1991

Issues

- tokenization
- differences in tagsets
  - Penn Treebank (various versions)
  - Universal Dependencies project

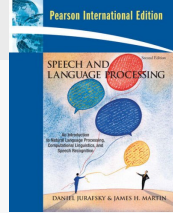


Part-of-Speech Tagging

- The process of assigning a POS marker to each input token
- It's a **disambiguation** task
  - *book that flight*
  - *hand me that book*
  - *Does that flight serve dinner*
  - *I thought that your flight was earlier*
- The goal of POS tagging is to **resolve** these ambiguities

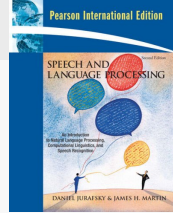
Types:		**WSJ**	**Brown**
Unambiguous	(1 tag)	44,432 (**86%**)	45,799 (**85%**)
Ambiguous	(2+ tags)	7,025 (**14%**)	8,050 (**15%**)
Tokens:			
Unambiguous	(1 tag)	577,421 (**45%**)	384,349 (**33%**)
Ambiguous	(2+ tags)	711,780 (**55%**)	786,646 (**67%**)

Figure 8.2 Tag ambiguity for word types in Brown and WSJ, using Treebank-3 (45-tag) tagging. Punctuation were treated as words, and words were kept in their original case.



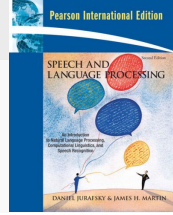
Example

- *That*, *back*, *down*, *put* and *set* are among the most frequent ambiguous words
  - *earnings growth took a back seat*
  - *a small building in the back*
  - *a clear majority of senators back the bill*
  - *Dave began to back toward the door*
  - *enable the country to buy back its debt*
  - *I was twenty-one back then*



Example

- *That*, *back*, *down*, *put* and *set* are among the most frequent ambiguous words
  - *earnings growth took a back/JJ seat*
  - *a small building in the back/NN*
  - *a clear majority of senators back/VBP the bill*
  - *Dave began to back/VB toward the door*
  - *enable the country to buy back/RP its debt*
  - *I was twenty-one back/RB then*
- However, not all tags are equally likely
- **Most Frequent Class Baseline**
  - Always compare a classifier against a baseline at least as good as the most frequent class baseline
  - most-frequent-tag-baseline on WSJ corpus: 92.34% accuracy
  - State-of-the-art accuracy: 97-98%



Approaches to POS Tagging

- Probabilistic methods like Hidden Markov Model
- HMM per-token accuracy on WSJ corpus around 96.5%
- However, even 3% per-token errors means 55-57% sentence-accuracy
  - a single bad mistake in a sentence can greatly throw off the usefulness of a tagger to downstream tasks such as dependency parsing
- Bidirectional approaches like CRF and BiLSTM (ca. 97.85%)
- Contextual word embeddings

POS Tagging (State of the art), ACL Wiki
Part-of-Speech Tagging from 97% to 100%:Is It Time for Some Linguistics?, C Manning, 2011

A1: RNNs for Sequence Labeling

Andrea Galassi

$\rightarrow$ Virtuale $\leftarrow$

A2: QA with Transformers on CoQA

Federico Ruggeri

$\rightarrow$ Virtuale $\leftarrow$

Guidelines for a Successful Completion of Assignments and Projects

Rule Number One

- When emailing us, always include **all** members of the teaching staff as recipients (To:) or in carbon copy (Cc:)
  - Paolo Torroni
  - Andrea Galassi
  - Federico Ruggeri

(it's easy to miss emails otherwise)

Coursework Submission and Presentation

- **What to submit**: a **single ZIP archive** that includes the assignemnt or project itself (python code, datasets if applicable, etc) and a report
- **How to submit**: on virtuale, using the assignment link or project submission link.
  - Only one submission per team, made by the speaker of the team.
- There is **no assignment presentation**, only submission followed by grading (via virtuale)
- **Project Presentation**: after the project report has been submitted on virtuale by the speaker of the team, **each** team member must sign up on **AlmaEsami** for the project presentation.
  - Presentations must last 15 mins tops
  - Booking of presentations to be confirmed based on evaluation of report and daily capacity (max 7 teams per day)
  - First date January 1; see AlmaEsami for more dates

Project Report

- Fixed structure
  - Abstract
  - Introduction
  - Background (only for project reports)
  - System description
  - Data (only for project reports)
  - Experimental setup and results
  - Discussion
  - Links to external resources (optional)
  - References
  - Appendices (optional)
- Assignments: 2-page reports; Projects: 8-page reports
- Template available

  → Virtuale ←

Coursework Evaluation

How are Projects Evaluated

- **Project**: up to 10 points
  - methodology
  - implementation
- **Report**: up to 4 points
  - clarity and quality of technical presentation
  - motivation of design choices
  - positioning with respect to relevant literature
  - discussion of experimental setup and results, including error analysis
  - discussion of limitations, alternatives, possible improvements
- **Discussion**: up to 4 points
  - may extend to any topics covered during the course
  - make sure each team member speaks, don't exceed allotted time, be clear and to the point:
    - what is your contribution?
    - what are significant results / observations?
    - why did you make the choices you made?

Success Criteria in Project Evaluation

- **Implementation**: well-documented and easily readable code
  - meaning: either self-explaining code, or containing enough comments to easily understand what's going on without having to struggle
- **Methodology**:
  - usage of clearly defined splits
  - demonstrating understanding of evaluation metrics and using the right evaluation metrics
  - carrying out good experimentation with pre-processing and machine learning models
  - definition of relevant/fair baselines for comparison
  - error analysis and/or quantitative analysis
  - analysis of model behaviour in interesting cases, if applicable

Things that Don't Matter

Things that have **no impact** on the evaluation

- Absolute model performance
- Performance relative to other groups
- Submission date

Things to Avoid

Caveats:

- Project submitted by teams composed of
  - less than 3 members or
  - more than 4 members

  will not be evaluated, unless different size agreed with teaching staff
- Incomplete projects - for example, without report - will not be evaluated.

Warning: if you submit coursework after February 2023, do **notify us by email**, otherwise we may not be aware you're waiting for our evaluation.

Cut-off Dates

- We will check submissions **starting one week before the discussion date** published on AlmaEsami
- Submissions uploaded later than 7 days ahead of a discussion date will be scheduled to the next discussion date

Standard Project

→ SemEval 2023 Task 4: ValueEval ←

Identification of Human Values behind Arguments

Identification of Human Values behind Arguments

- **Objective**: Given a textual argument and a human value category, classify whether or not the argument draws on that category
- Based on Kiesel et al, Identifying the Human Values behind Arguments. ACL 2022
  - It's a paper with code
- Arguments are given as **premise text, conclusion text, and binary stance** of the premise to the conclusion ("in favor of" or "against").
- 20 **value categories** compiled from the social science literature
- It's **your choice** to focus on one, a subset, or all values in arguments.

→ Let's look at the website ←

Resources

- Touché at SemEval 2023 shared task website
- Huggingface portal for useful models
- Scientific articles from relevant conferences and journals
  - IJCAI, AAAI, ECAI, ACL, EMNLP, NAACL, EACL, COLING, LREC, TACL, COLI, TOIT, AIJ, JAIR, ...
- Whatever resource you are using, make proper and clear reference to it in the report

3CFU Project Work Extensions

- If you also have the 3CFU Project Work in your study plan, you can
  - either do your PW independently of your NLP project
  - or work at a larger NLP project, which will count for NLP+PW (9 CFUs total, of which 6 graded will contribute to your CGPA, 3 pass/fail on PW won't)
- This can be done with the standard project or with the custom project (no difference)
- Either way, whether you are going for an independent PW, or for a single big NLP project+PW, **before you start working on your PW or extended project** you should **check with the teaching staff** and have their approval that your intended work is indeed worth +3CFU
  - Send us an email, we'll set up a meeting to discuss

Custom Project

Choose Whatever You Like

First of all, we encourage you to work on something that you like and that you think it can be useful to you

- You can take an existing problem, with corresponding data, and address it applying processing techniques and training models for it
- Or you can work more on the side of the data, creating a novel dataset and applying "standard" techniques on that
- You can do anything in between, by curating a dataset from existing data and applying something a bit more than standard

Check NLP Workshops on Themes You Like

- Many workshops in the community of Computational Linguistics and Information Retrieval offer interesting challenges and shared tasks
- You can approach novel challenges but also challenges that have been done in the past
- Look for pointers on our Language Technologies Lab website

CLEF Conference

- CLEF is at the intersection between NLP and Information Retrieval
- Focus: **Multilingualism** and **Multimodal NLP**
- Has 10+ workshops called "labs" which promote multiple tasks
- Check the this year's tasks
- Some examples:

BioASQ	Question Answering on biomedical topics
Check That!	fact-checking and similarity between claims
PAN	stylometry and digital forensics
EXIST	sexism and abusive language identification

SemEval

- Collection of Workshops on **Semantic Evaluation**
- Each year there are different tasks that span across many domains
- Check this year's tasks
- Some examples:

Task 2	Multilingual Complex Named Entity Recognition
Task 3	Detecting the Category, the Framing, and the Persuasion Techniques in Online News in a Multi-lingual Setup
Task 7	Multi-Evidence Natural Language Inference for Clinical Trial DataTask
Task 9	Multilingual Tweet Intimacy Analysis
Task 10	Towards Explainable Detection of Online Sexism

EvalITA

- NLP tasks on **Italian Language**
- Check this year's tasks
- Some examples:

  Ghigliottin-AI Artificial Players for the Language Game "La Ghigliottina"

  CHANGE-IT Style Transfer

  ATE_ABSITA Aspect Term Extraction and Aspect-Based Sentiment Analysis

  HaSpeeDe Hate Speech Detection

Questions?