CLiPS
Computational Linguistics & Psycholinguistics
University of Antwerp

Profiling Personality of Social Media Authors

Walter Daelemans

CLiPS Computational Linguistics Group

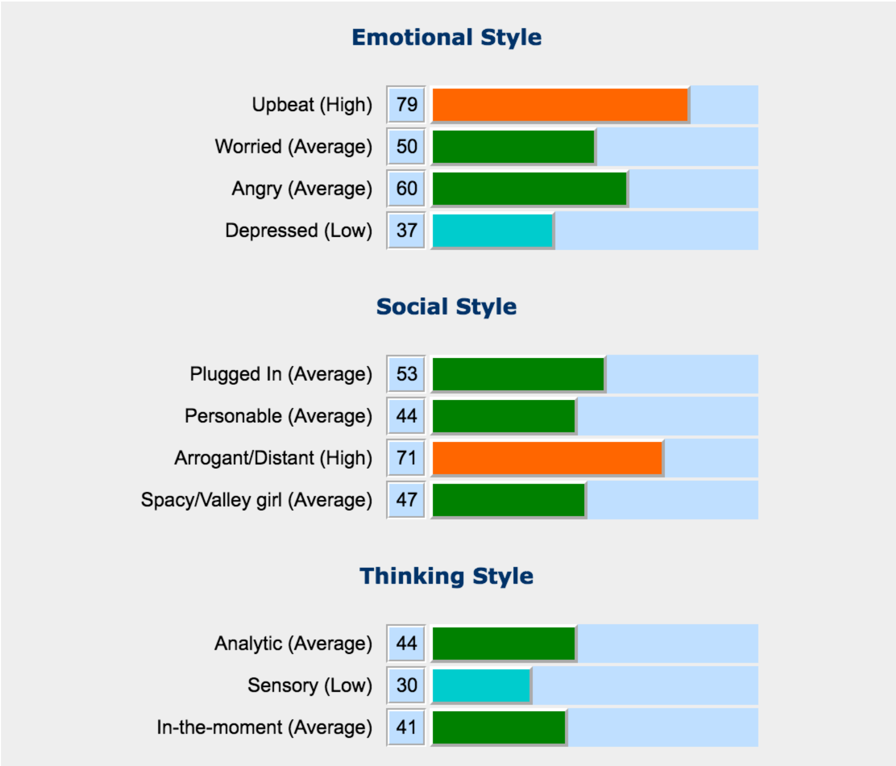
walter.daelemans@uantwerpen.be

SLSP 2016, Pilsen

October 11, 2016

www.analyzewords.com (Pennebaker)

See also: https://personality-insights-livedemo.mybluemix.net/ (Watson)



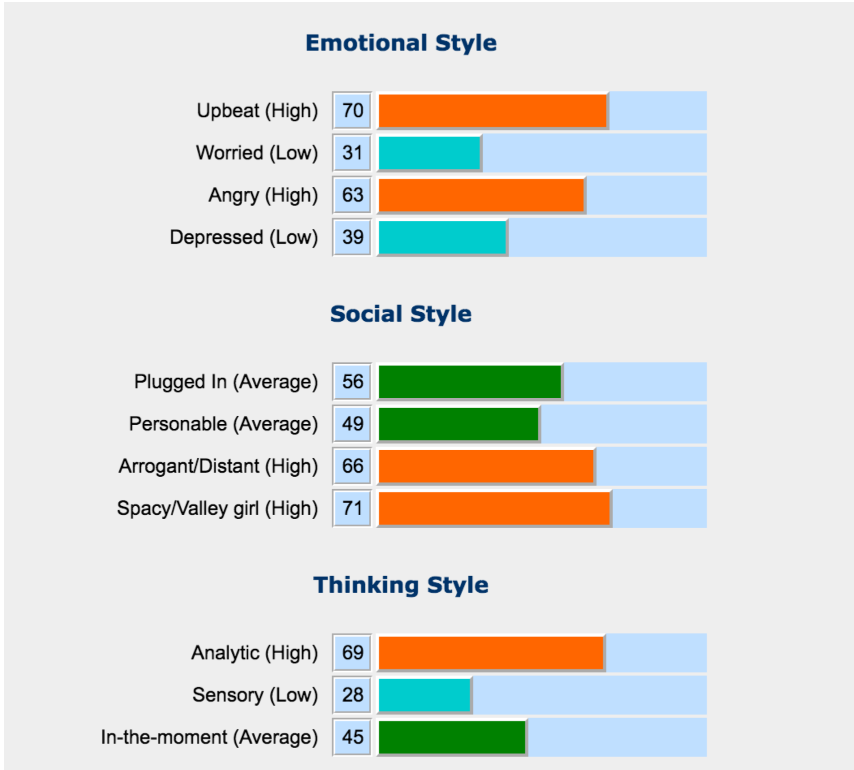
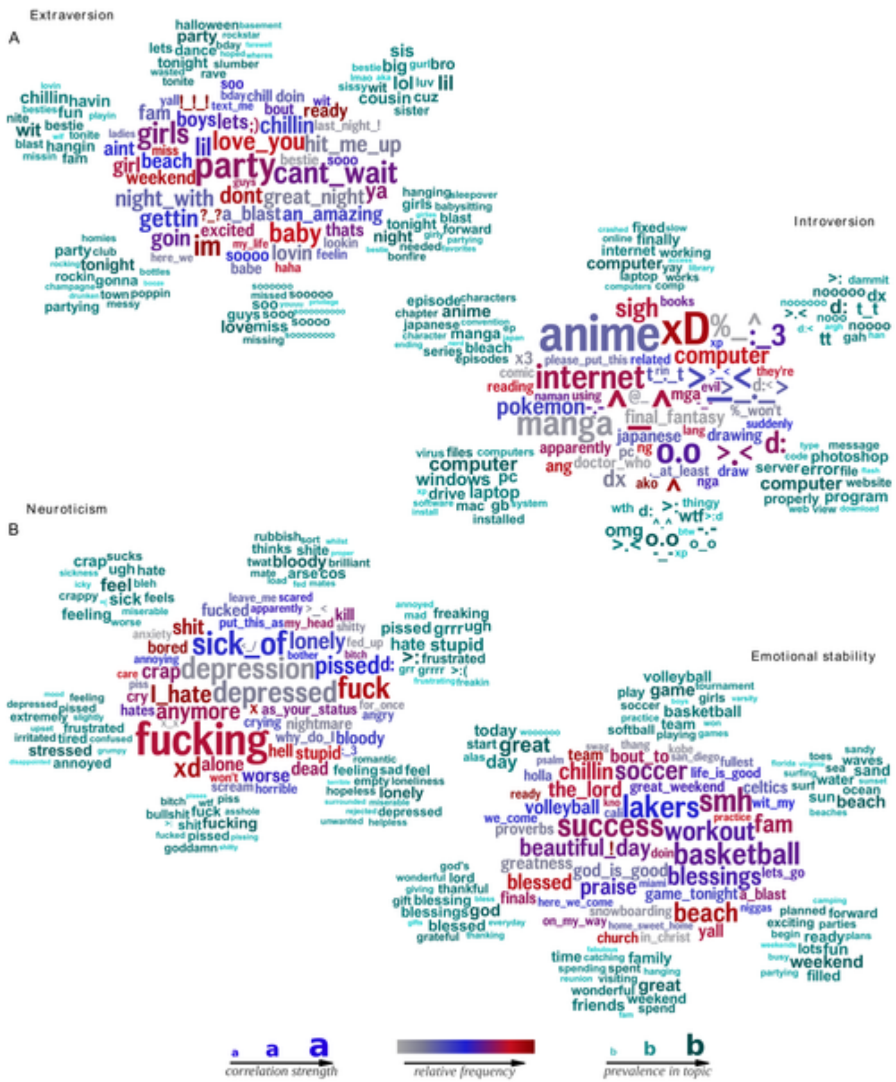


Figure 6. Words, phrases, and topics most distinguishing extraversion from introversion and neuroticism from emotional stability.



Schwartz HA, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones SM, et al. (2013) Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. PLoS ONE 8(9): e73791. doi:10.1371/journal.pone.0073791
http://journals.plos.org/plosone/article?id=info:doi/10.1371/journal.pone.0073791

What we don't want to do

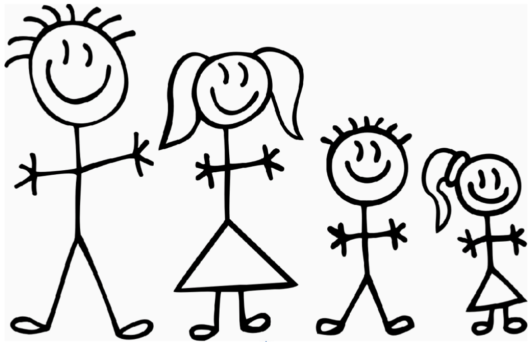
- Horoscopes
  - E.g. the profiling of Pennebaker / IBM personality
    - We need gold standard data
- Content-based assignment
  - E.g. emotionally stable American people talk more about “sports”, “vacation”, “beach”, “church”, or “team”
    - It’s not style and can be faked easily

Contents

- Profiling and Computational Stylometry
  - Methods and Issues
  - Social Media Language
- Personality Profiling in Social Media
  - Applications
  - State of the Art
    - PAN
    - CGI and Twisty corpora

What is Profiling?

Individual Variation | **Language Variation**



Sociolinguistics

Spelling, punctuation, lexical choice, sentence structure, themes, tone, discourse structure, …

Profiling

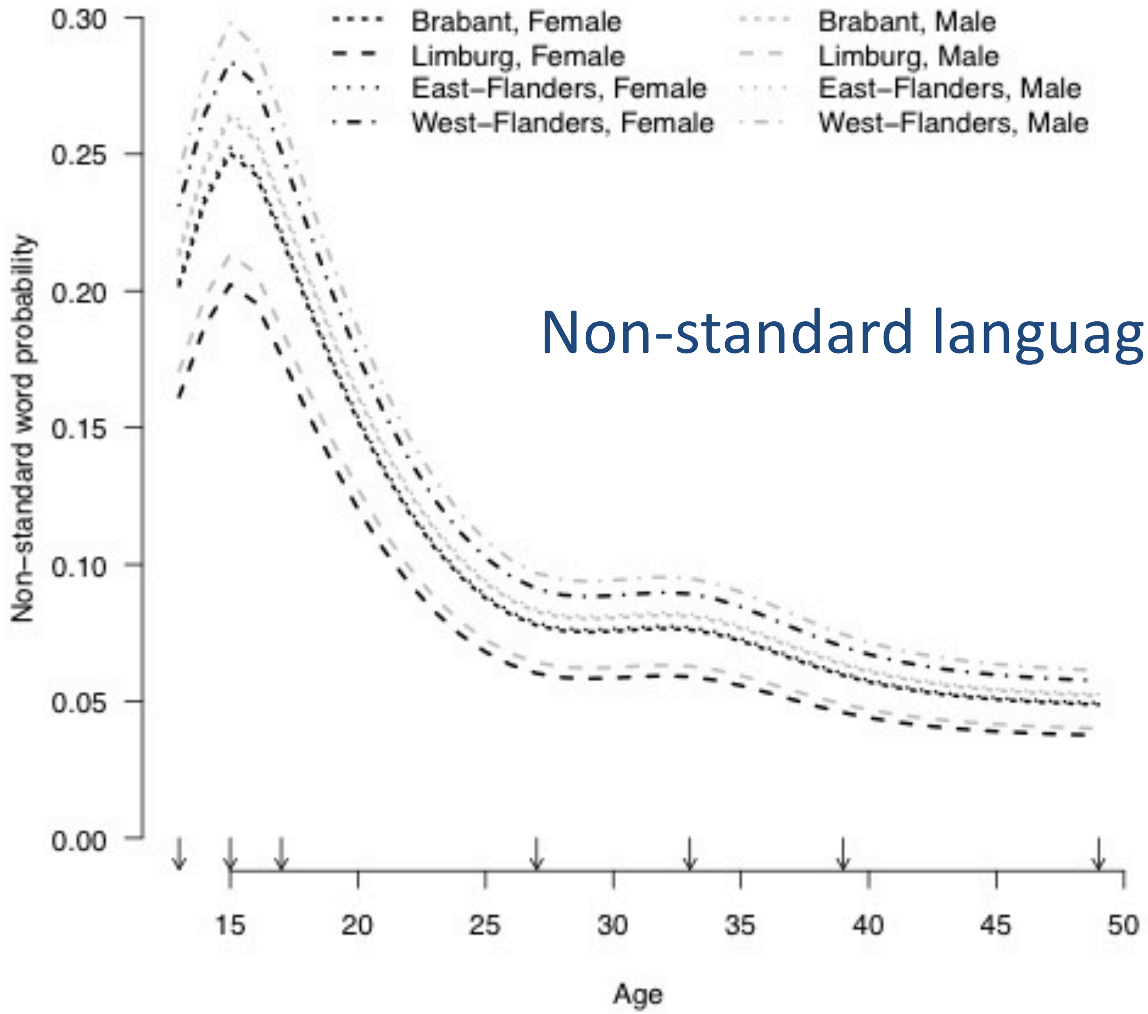
$$\mathrm{P}(i|l) = \frac{P(l|i)P(l)}{P(i)}$$

Bayes rule (1763)

Social Media language

- *Everybody* "writes" all the time now, not only professional writers
- Huge increase in subjective language (opinions) as opposed to objective, factual language
- Open sharing of private information (metadata)
  - Magnificent source of data for computational stylometry!

Computational Sociolinguistics!



Non-standard language use in chat

Computational Stylometry

- Writing style: A *combination* of *invariant* and *unconscious* decisions in language production at all linguistic levels (discourse, syntactic structures, lexical choice, …) *associated* with specific authors or author traits:
  - Age, gender, education level, native language, **personality**, emotional state, mental health, region, deception, ideology, political conviction, religious beliefs, sexual preferences, …

Basic research questions

- Does an idiolect / stylome exist and (how) can it be measured?
- Can we handle genre, register & topic interference?
- Can we handle within-profile interference?
- Is detection robust? (adversarial stylometry)
- Can we handle dynamic aspects? (language change over time with age, illness, language input, context, …)
- Can we explain why / how trained models work?

Method: Text Categorization

Document → Document Representation

Feature construction

Superficial
- token n-grams
- character n-grams
- punctuation patterns

Lexical
- lemma n-grams
- readability features
- lexical richness

Syntactic
- Function word patterns
- Part of speech n-grams
- Syntactic structures

Semantic
- word sense patterns
- semantic dictionaries
- distributed vectors
- word embeddings
- semantic role patterns

Discourse
- rhetorical structures
- discourse marker patterns

Input from linguistic theory

Method: Text Categorization

Document Representation → Feature Vector

Feature Selection

Superficial
- token n-grams
- character n-grams
- punctuation patterns

Lexical
- lemma n-grams
- readability features
- lexical richness

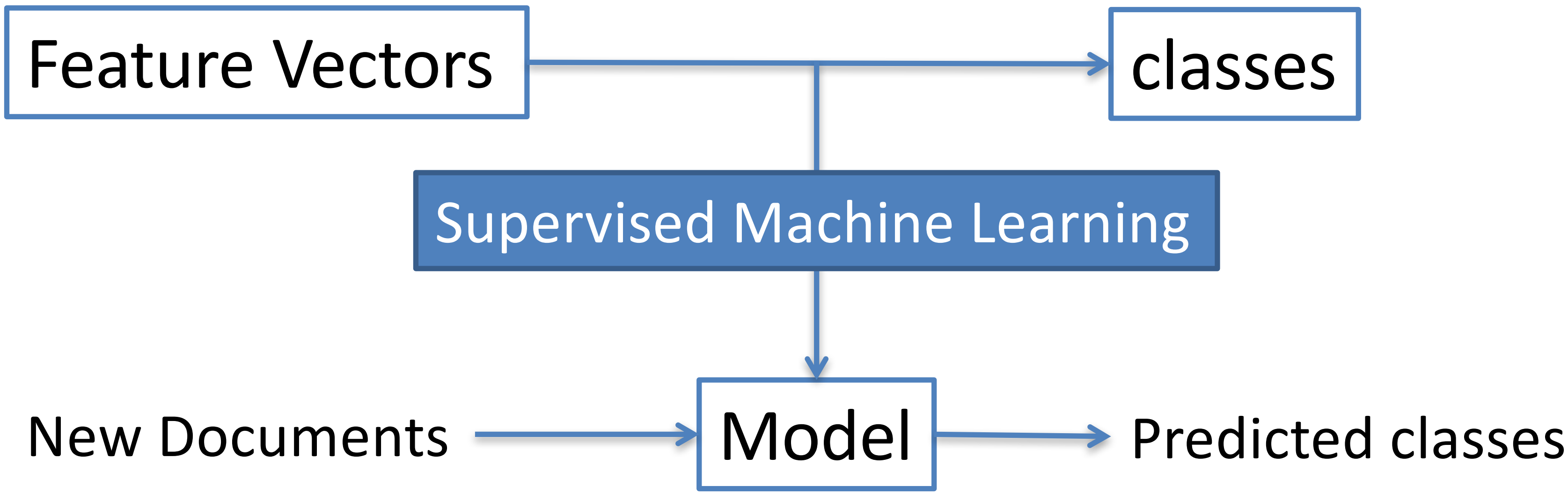
Syntactic
- Function word patterns
- Part of speech n-grams
- Syntactic structures

Semantic
- word sense patterns
- semantic dictionaries
- distributed vectors
- word embeddings
- semantic role patterns

Discourse
- rhetorical structures
- discourse marker patterns

Feedback to linguistic theory

Method: Text Categorization



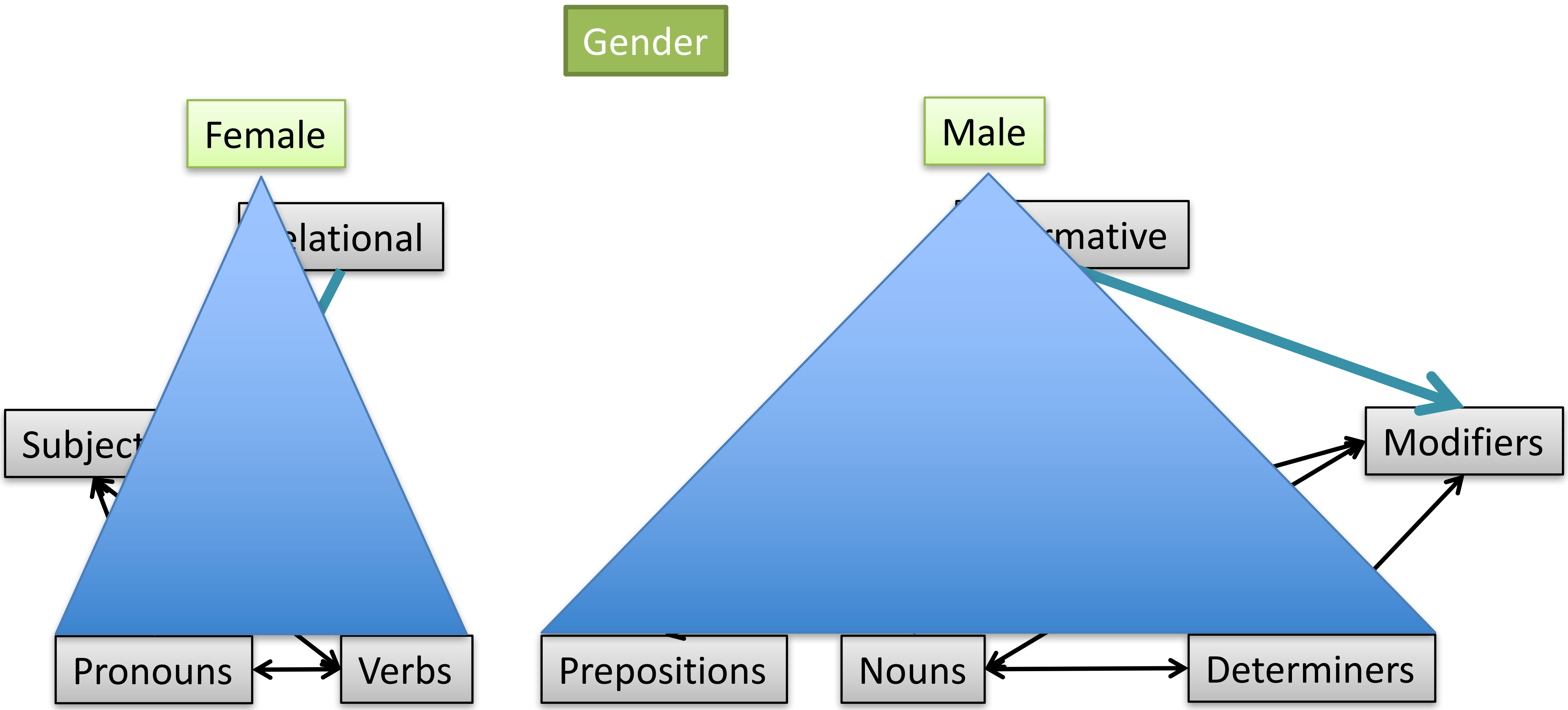
- Objective evaluation on the basis of “gold standard data” and evaluation metrics

Feedback to linguistic theory

Explanation in Stylometry

- Quantitative evaluations and lists of important features don’t tell us a lot about how it works
- *Why* can age, gender, personality, authorship etc. be determined with a particular set of linguistic features?
- Landmark: gender in fiction and non-fiction
  - Koppel & Argamon et al. 2002

Explanation in Stylometry



Personality

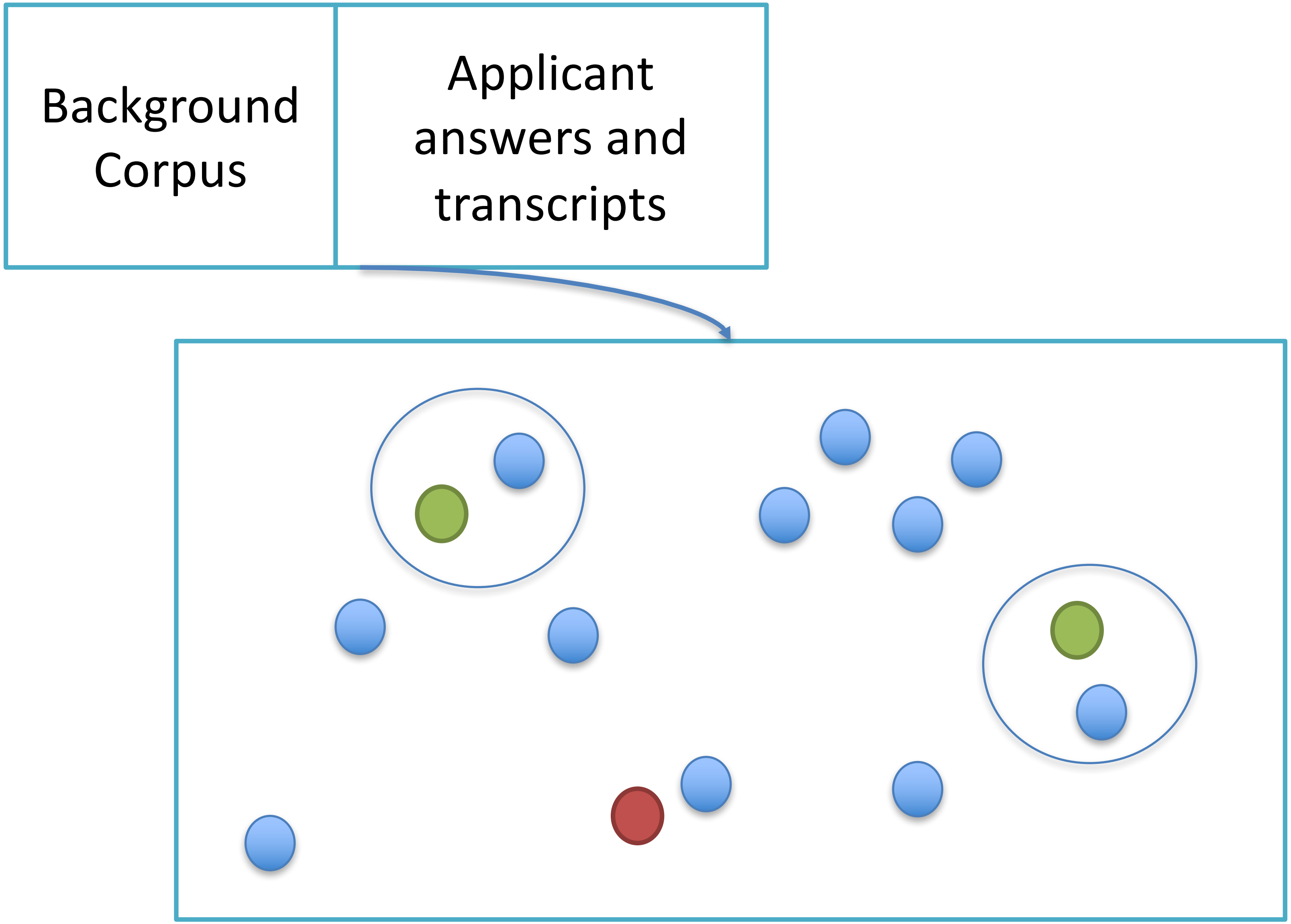
- Set of mental traits (types) that
  - Explain and predict patterns of thought, feeling and behavior
  - Remain relatively stable over time and context
- Source of variation between people that
  - Influences success in relations, jobs, studies, …
  - Explains 35% of variance in life satisfaction
    - Compare: income (4%), employment (4%), marital status (1% to 4%)
- Detection interacts with that of deception, opinion, emotion, …

Applications of Personality Detection

- Targeted advertising
- Adaptive interfaces and robot behavior
  - music, style, tone, color, …
- Psychological diagnosis, forensics
  - psychopaths, mass murderers, bullies, victims
  - depression, suicidal tendencies
- Human resources management
- Research in Literary science, social psychology, sociolinguistics, …

Example: Human Resources

- Situation: thousands of applicants for 1 job
- Method
  - Collect written answers to questions and transcripts of audio messages
  - Deep learning of word embeddings on background text
  - Deep learning of “applicant embeddings” on collected text
  - Similarly for “anchors” (model answers, good candidates)
  - Nearest neighbor ranking



Applicant Embedding Map

Data collection

- Source
  - Social media
    - Twitter, Facebook statuses
  - Essays
- Method
  - Self-reporting
    - questionnaire
  - Expert annotation (consensus)
  - Distant supervision
    - Patterns extracted by experts from questionnaire questions (Neuman & Cohen) or from correlation studies (Celli)
- Class system (Classification of Regression)
  - MBTI (Myers-Briggs Type Indicator)
  - FFM (Five Factor Model, Big Five)

Big Five (FFM)

- Extraversion (vs. Introversion)
  - Sociable, active, energetic, positive
- Neuroticism
  - Sad, anxious, tense
- Agreeableness
  - Altruism, trust, modesty, sympathy
- Conscientiousness
  - Goal-oriented, organized, in control
- Openness
  - Originality, breadth and depth of mental life and experience

Meyers-Briggs preferences

- **I**ntroversion & **E**xtraversion
- i**N**tuition & **S**ensing
- **F**eeling & **T**hinking
- **J**udging & **P**erceiving

- Leads to 16 types: ENTJ (1.8%) … ESFJ (12.3%)
- Validity and reliability have been questioned

A typical sample of our students

28 ESFJ (provider)
23 ENFJ (teacher)
16 ISFJ (protector)
15 INTJ (mastermind)
15 INFJ
9 ENFP
8 ISTJ
8 ENTJ

6 ESFP
4 ISFP
4 INFP
4 ESTJ
3 INTP (architect)
1 ESTP (promoter)
1 ENTP (inventor)
0 ISTP (crafter)

Feature Correlations

- Extraverts (as opposed to introverts)
  - Produce more language (verbosity)
  - Smaller vocabulary (TTR), fewer hapaxes
  - Fewer negative emotion words
  - More positive emotion words
  - Fewer hedges (confidence)
  - More agreement and compliments
  - Fewer negation and causation
  - Less concrete
  - Less formal, more contextualised / relational
    - More pronouns, verbs, adverbs, interjections
    - More present tense
    - Fewer numbers, less quantification
    - Fewer negation and causation words

Feature Correlations

- Neurotics
  - More “I”
  - More negative emotion words, fewer positive emotion words
  - More concrete and frequent words
- Agreeable
  - More positive emotion words, fewer negative emotion words
  - Fewer determiners
- Conscientious
  - Fewer negations
  - Fewer negative emotion words
  - Fewer hedges (?)
- Openness
  - Longer words
  - More hedges
  - Fewer “I” and present tense

PAN 2015

- Personality as part of profiling (also age and gender)
- Twitter
- 5 numeric FFM values
  - Regression (one classifier for each factor)
  - Evaluation: RMSE
- English (194 users), Spanish (140), Italian (50), Dutch (44)
- 22 teams

Overview of the 3rd Author Profiling Task at PAN 2015
F Rangel, P Rosso, M Potthast, B Stein, W Daelemans - CLEF, 2015
http://pan.webis.de/clef15/pan15-web/author-profiling.html

Document Representation

- Preprocessing
  - Remove HTML codes, hashtags, URLs, mentions, …
- Features
  - character n-grams
  - word n-grams
  - tf-idf n-grams
  - syntactic n-grams (lemma, pos, relations, …)
  - stylistic features
    - punctuation, case
    - emoticons
    - word, sentence length, verbosity
    - character flooding
  - topic modeling (LSA)
  - family vocabulary, LIWC, MRC
  - frequent terms, discriminative words, Nes, other vocabularies ...
  - second order features (relationships among terms, documents, profiles)

Document Representation

Álvarez-Carmona et al. (2015)

- Preprocessing
  - Remove HTML codes, hashtags, URLs, mentions, …
- Features
  - character n-grams
  - word n-grams
  - **tf-idf n-grams**
  - syntactic n-grams (lemma, pos, relations, …)
  - **stylistic features**
    - punctuation, case
    - emoticons
    - word, sentence length, **verbosity**
    - character flooding
  - **topic modeling (LSA)**
  - family vocabulary, LIWC, MRC
  - frequent terms, discriminative words, Nes, other vocabularies ...
  - **second order features (relationships among terms, documents, profiles)**

Machine Learning Method

- SVMs
- Decision Trees (Random Forests)
- Bagging
- Linear Discriminant Analysis
- Stochastic Gradient Descent
- Linear, Logistic, Ridge for Regression
- Distance-based approaches

Best Results

	O	C	E	A	N
English	.12	.11	.14	.13	.20
Spanish	.11	.10	.13	.10	.16
Italian	.10	.11	.07	.05	.16
Dutch	.04	.06	.08	.00	.06

Problem

- Same features are informative for different author aspects (e.g. personality & gender)
  - Women ~ extraverts
    - positive and negative emotion words, pronoun use, …
  - Joint learning, ensemble methods, …
- Main cause of overfitting
- Large **reference corpus** needed of authors with various traits of interest
  - **Sample stratification**
    - Example: CLiPS Stylometry Investigation (CSI) corpus

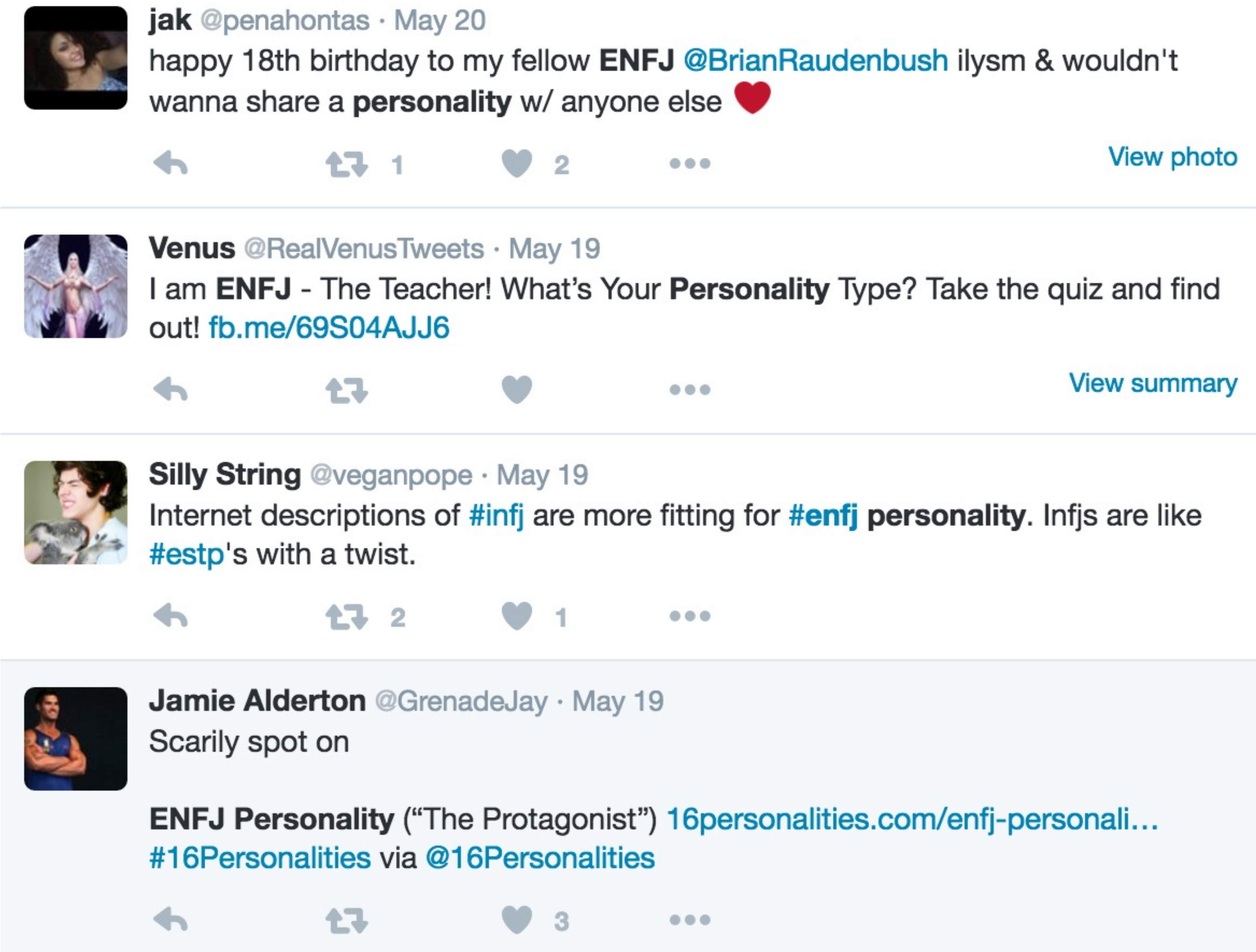
CLiPS CSI Corpus (with Ben Verhoeven)

- http://www.clips.uantwerpen.be/datasets/csi-corpus
- Corpus with two genres: essays and reviews (truthful and deceptive)
  - Dutch native speakers, students of language and literature
- Meta-data
  - Age, gender, sexual orientation, region of origin, personality profile (Big Five & MBTI)
- Yearly expansion
  - Currently: 1800 documents; 660 authors; 770,000 words

- Successor of Personae Corpus (2008, with Kim Luyckx)

Mining personality data from Twitter (with Verhoeven and Plank)

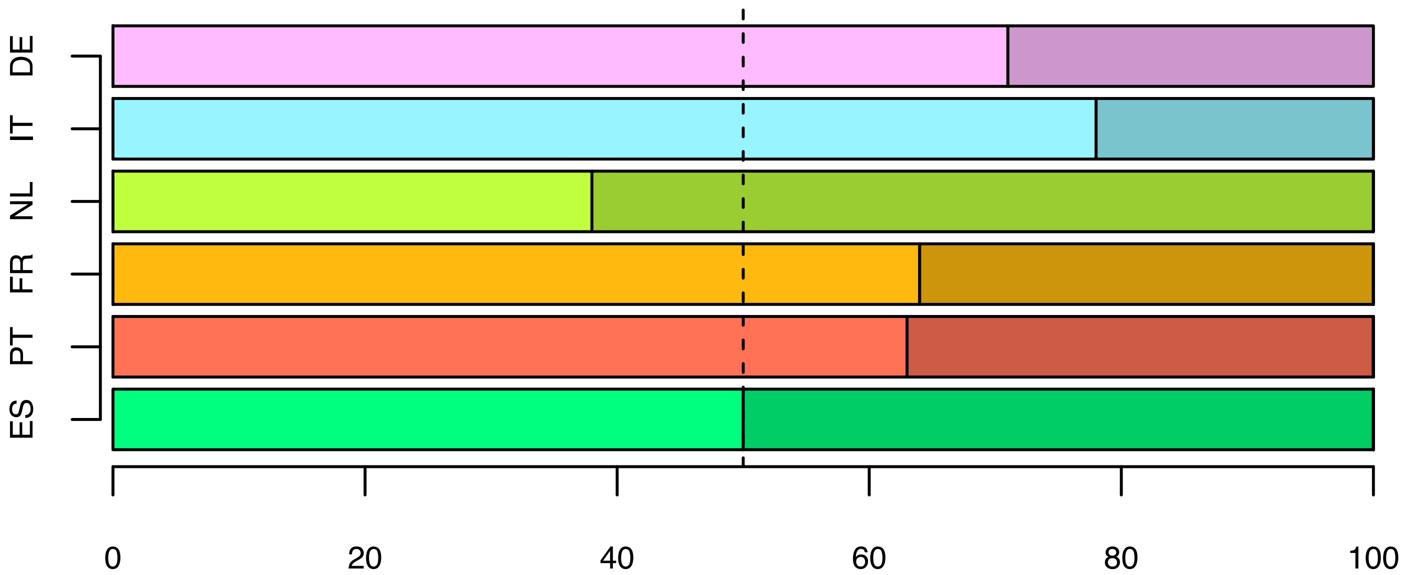
- People tweet about their own personality



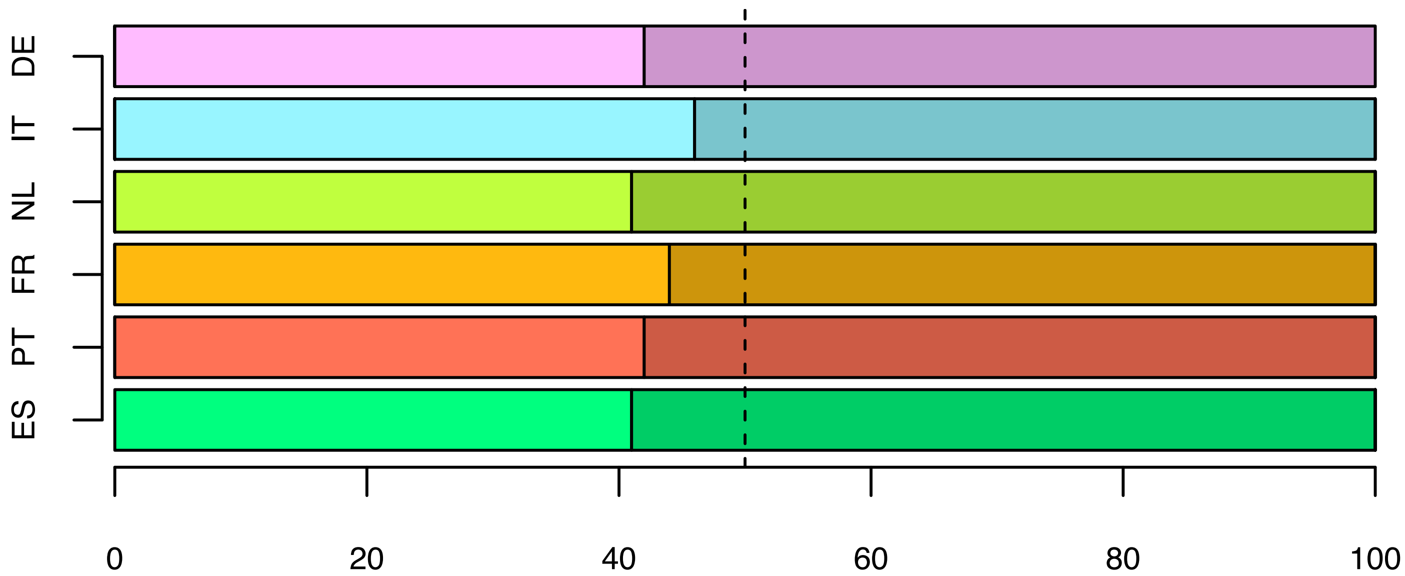
TwiSty Corpus approach

- Search users with a tweet mentioning their MBTI personality type (keyword-based) + gender
  - manual checking
- Fetch recent tweets from those users (~ 2000 average)
  - apply language identification for selecting clean language sample
- ES, PT, FR, NL, IT, DE
- Sensationalism bias?
- TwiSty is available from http://www.clips.uantwerpen.be/datasets

Some corpus properties



I versus E



T versus F

Prediction Experiment

- Linear SVC, word and character n-grams

Lang	Task	WRB	MAJ	P	R	F
DE 387	I–E	60.22	72.61	71.43	73.1	72.27
	S–N	71.03	82.43	67.95	82.43	74.49
	T–F	51.16	57.62	58.38	59.69	**59.03**
	J–P	53.68	63.57	60.27	63.82	61.99
	Gender	50.28	53.75	77.72	77.52	**77.62**
IT 443	I–E	65.54	77.88	76.42	79.23	77.78
	S–N	75.60	85.78	73.58	85.78	79.21
	T–F	50.31	53.95	51.66	52.60	52.13
	J–P	50.19	53.05	46.63	47.40	*47.01*
	Gender	54.78	65.46	73.90	72.69	**73.29**
NL 920	I–E	53.02	62.28	61.82	64.02	**62.90**
	S–N	57.66	69.57	69.39	71.63	**70.49**
	T–F	51.47	58.59	59.26	60.65	**59.95**
	J–P	52.00	60.00	56.50	59.57	57.99
	Gender	50.04	51.41	82.62	82.61	**82.61**

FR 1,250	I–E	54.77	65.44	65.35	67.68	**66.49**
	S–N	68.00	80.00	77.60	80.24	78.90
	T–F	50.65	55.68	57.88	58.56	**58.22**
	J–P	52.13	60.32	55.06	58.64	56.79
	Gender	51.84	59.60	83.77	83.84	**83.80**
PT 3,867	I–E	53.36	62.97	66.06	67.34	**66.69**
	S–N	63.60	76.08	71.02	75.98	73.42
	T–F	51.27	57.98	61.23	62.01	**61.62**
	J–P	50.87	56.61	56.10	56.97	56.53
	Gender	52.15	60.36	87.54	87.56	**87.55**
ES 9,445	I–E	50.00	50.49	61.09	61.09	**61.09**
	S–N	55.42	66.47	60.23	62.91	61.54
	T–F	51.63	59.04	59.35	60.12	**59.73**
	J–P	51.53	58.75	55.60	56.56	56.08
	Gender	51.00	57.06	87.61	87.63	**87.62**

Cognitive-biological Approach (Yair Neuman)

- “Personality” categorization originates from
  - Threat / Trust management
  - Risk assessment process
    - Fight, flight, avoid
  - (+ complexity of abstraction and inference unique to humans)
- in the context of Interpersonal relations

Conclusions

- Computational personality
  - Promising for trend-based applications but not good enough yet for individual profiling
  - Text categorization model is not sufficient
    - Hard to obtain sufficient reliable annotated data
    - Unsupervised models? Clever mining?
    - threat - trust model?
  - Computational theory of writing style needs a large, preferably multilingual, but in any case balanced corpus

Stylometry researchers @ CLiPS:

Questions?

CLiPS
Computational Linguistics & Psycholinguistics
University of Antwerp