Orientation

Alexander Fraser & Liane Guillou
{fraser,liane}@cis.uni-muenchen.de

CIS, Ludwig-Maximilians-Universität München

Computational Morphology and Electronic Dictionaries
SoSe 2016
2016-04-11

Outline

Course Information

General information

- Lecture (Vorlesung): Monday 16:15 – 17:45 in room L155 (or occasionally computer pool)
- Exercise (Übung): Wednesday 10:15 – 11:45 in room L155 (will often be in computer pool)
- There will not be a strict separation of lectures and exercises
- Schedule and lecture slides posted on web page (see Fraser homepage, Google: fraser CIS)

Change This Semester

Computational Morphology and Electronic Dictionaries in the Past:

- **Primarily Linguistic Theory**
  - Strong focus on linguistics behind morphological processes
  - Some motivation through applications (particularly last semester, Springmann)
  - No programming (this becomes a problem later, if you want to do the CL Masters!)
  - No finite-state morphological analysis

Course information

Contents and goals of this course

This course will look at morphology from 2 perspectives:

- **From a computational side**
  - Understanding the challenges of modeling morphological phenomena computationally
  - Connections with NLP applications
  - Understanding the background behind regular expressions and finite automata
  - Focus on PERL regular expressions, some PERL programming
  - Working with a Finite-State Toolkit

- **From a theoretical side**
  - Basic concepts in morphological theory
  - Understanding the challenges of the theoretical modeling of morphological phenomena
  - Inflection, derivation and compounding

Course instructors

Fraser and Guillou

- **Dr. Alexander Fraser**
  - Fraser: will be teaching mostly computational lectures in German (using English slides)
  - One of his research foci is applying computational morphology in machine translation (e.g., from English to German)
  - Dr. Fraser is a permanent staff member at CIS (and coordinator of the Masters program), leads three large research projects

- **Dr. Liane Guillou**
  - Guillou: will be teaching mostly morphological theory in English
  - One of her research foci is discourse phenomena in machine translation, including morphologically marked phenomena (e.g., pronouns)
  - Dr. Guillou has just joined CIS as a post-doc in Fraser's group

- **Tutor: Luisa Berlanda**
  - Luisa Berlanda will be the tutor for this course

Course Material

The course material is mainly based on the books:

- **Theoretical background**

  Haspelmath, M. & Sims, A.D. (2010):
  Understanding Morphology, 2nd edition
- Chapters 1 to 4
- REQUIRED

- **Finite State Morphology**

  Beesly, K.R. & Karttunen, L. (2003): *Finite State Morphology*

  `http://www.fsmbook.com`
- OPTIONAL

Course Requirements

- To pass this course ...
  - Exercises and assignments
  - Regular attendance
  - Course project: implementation of a small project including extensive documentation; presentation
    * Roughly last 5-6 weeks of semester
    * Programming and data analysis intensive
    * Short presentation

Who should take this course

C++ versus Morphology

- **Higher Programming (C++)**
  - Prefer programming to looking at linguistic data
  - Important: C++ is required for some obligatory Masters courses
- **Morphology**
  - Prefer looking at linguistic data, ready to do some programming
  - Maybe you already know C++ and/or find the Higher Programming more basic than a different course in CS
  - Interested in working with linguistic tools such as morphological analysis, POS-tagging of morphologically rich languages like most Germanics, Slavics, etc.
- **Or take both!**
  - However, you will need to commit to getting a grade in just one course (in previous semesters students waited to see how they were doing, this will not work this semester)

Schedule over next two weeks

Schedule:

- Lecture, today: orientation, very brief introduction to morphology
- Exercise, Wednesday April 13th: attendance not required, individual advising/issues, etc.
- Lecture, Monday, April 18th: may be cancelled (with later make-up), CHECK COURSE WEB PAGE!
- Exercise, Wednesday April 20th: will NOT be cancelled, important that you attend if you will want a project later (= a grade in this class)
- Reading Assignment: **Read Chapter 1 of Haspelmath and Sims by Monday April 25th**

Questions?

Any questions about logistics, etc., before I briefly introduce morphology?

Outline

Acknowledgements

Some of the content of this lecture is based on previous lectures
by Marion Weller, Boris Haselbach, Özlem Çetinoğlu and Cerstin Mahlow.

Introduction

Words, words, words ...

- Words in natural languages encode many pieces of information
- What is the meaning of a word?
- How do words in a sentence interact with each other?
  - Subject/Verb agreement
  - Adjective/Noun agreement
  - ...
- What lexical category does a word belong to?
  - Noun (N)
  - Verb (V)
  - Adjective (A/ADJ)
  - ...
- What can we say about the internal structure of a word?
  - Determine the parts a complex word is composed of
  - Specify morphological features such as *number*, *gender*, *tense*, ...

Introduction

Internal structure of words: example

- **English**

  I am swim-m-ing

  - We know the meaning of *(to) swim*
  - -ing: marks the progressive form
  - Why the extra m?

- **Turkish**

  Ben yüz-üyor-um
  I.Nom *swim*-Prog-1P.Sg

  - yüz means 'swim'
  - -üyor corresponds to English *-ing*
  - -um indicates the person

⇒ Inflected Turkish verb contains more information

Introduction

Morphological processes

- **Inflection**
  Modification of a word to express different grammatical categories (*number*, *gender*, *tense*, ...)
  - *dog* → *dogs*
  - *write* → *writes*
- **Derivation**
  Process of forming a new word using an existing one
  - *happy* → *happiness*
  - *essen* → *essbar*
- **Compounding**
  Creating a new word containing two or more pre-existing words
  - *Apfel+Kuchen* → *Apfelkuchen*
  - *Donau+Dampf+Schiff+Fahrt+Kapitän+Mütze* → *Donaudampfschifffahrtskapitänsmütze*

Introduction

Two challenges

- **Morphosyntax (Morphotactics)**
- Words are composed of smaller units (morphemes)
- When combining morphemes, certain rules/conditions need to be fulfilled

  piti-less-ness
  *piti-ness-less

- **Pholonogical/Orthographical Alternations**
- The realization of a morpheme might vary depending on its context ($\rightarrow$ allomorph: variation of a morpheme)

  pity $\rightarrow$ piti in pitilessness
  die $\rightarrow$ dy in dying
  swim $\rightarrow$ swimm in swimming

Introduction

Why is morphology important?

- Many NLP applications need to extract the information encoded in complex words
- Rich morphology leads to data sparsity
  English: blue → German: blau, blaues, blaue, blauen, blauem, blauer
- **Syntactic Parsing**
  To analyze sentence structure, a syntactic parser needs information about:
  - subject-verb agreement
  - adjective-noun agreement, ...
- **Information retrieval**
  Better generalization when working on lemmatized forms
- **Machine translation**
  Need to analyze the words on the source-side and generate words with specific morphological features in the target language (e.g., gender of articles, case of noun-phrases, etc...)

Reminder: Schedule over next two weeks

Schedule:

- Lecture, today: orientation, very brief introduction to morphology
- Exercise, Wednesday April 13th: attendance not required, individual advising/issues, etc.
- Lecture, Monday, April 18th: may be cancelled (with later make-up), CHECK COURSE WEB PAGE!
- Exercise, Wednesday April 20th: will NOT be cancelled, important that you attend if you will want a project later (= a grade in this class)
- Reading Assignment: **Read Chapter 1 of Haspelmath and Sims by Monday April 25th**

Thank you for your attention.