Towards an XFST and XLE Grammar Implementation of Lithuanian

Universität Konstanz

D. Cavar (EMU & ILIT), J. Bacyte, A. Leiderer, M. Pollich, V. Rehberg, B. Rosemann (University of Konstanz)

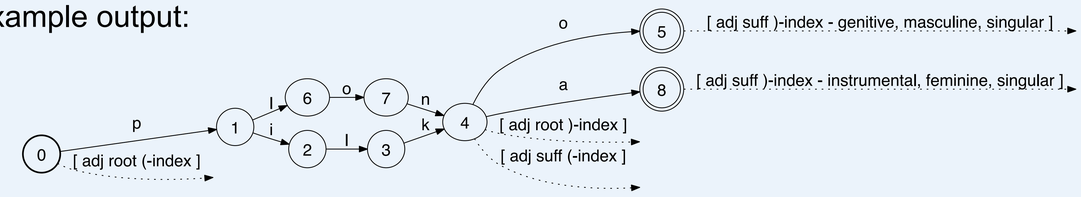
Goals
- No Lithuanian morphology or syntactic parser existent → First-time creation
- Realization of a Finite State Morphology using the Lexc-formalism and the Xerox XFST Compiler
- Definition of a LFG-based grammar for the Xerox Linguistic Environment (XLE)

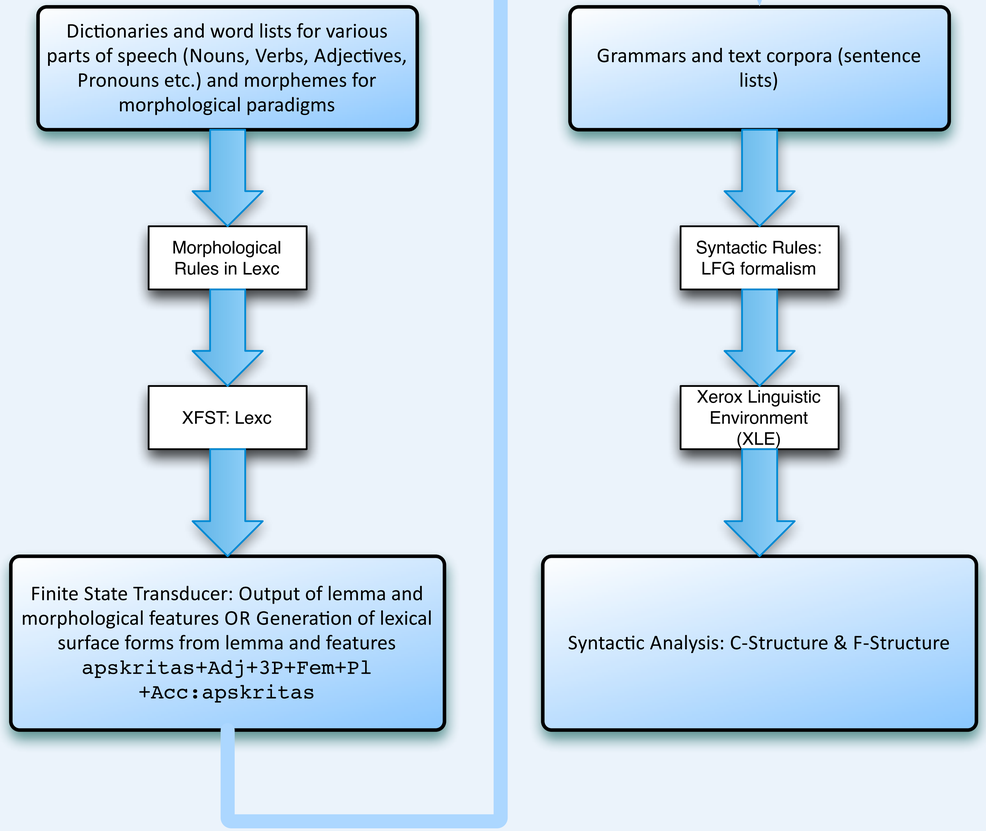
Lithuanian
- 3.2 million speakers
- Indo-European language: one of the last two Baltic languages
- 32 characters (Latin alphabet plus diacritics)
- Inflecting language
- Morpho-syntactic properties:
  - 7 cases (Nominative, Genitive, Dative, Accusative, Instrumental, Locative, Vocative),
  - 3 genders (Masculine, Feminine, Neuter)
  - 2 numbers (Singular, Plural) (partially Dual)

Morphology in Lexc → XFST Compiler
(Beesley and Karttunen 2003)

About 5000 morphemes
- Content words: Nouns, verbs, adjectives, adverbs
- Function words: Pronouns, auxiliaries, prepositions demonstratives, quantifiers, numerals
- Proper first and last names
- Inflectional paradigms and derivational rules
- Verb-valencies and subcategorisation frames.
- Extension via semantic features of stems possible

Resulting Finite State Transducer
Number states: 3656
Number of transitions: 7549
Example output:





Evaluation of Morphology
- Reading and tokenizing online newspapers using a python script
- Automatic entering of tokens in Finite State Morphology

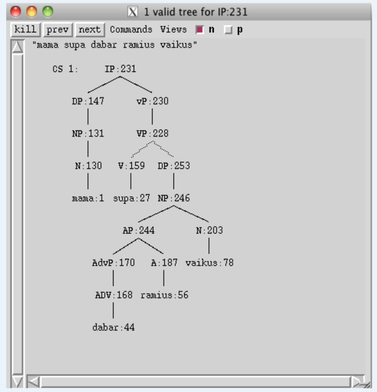
→ Extensive recognition of listed words in all inflectional forms
→ However, only fractional amount of all words listed at the moment

Grammar and Parser
Lexical Functional Grammar (LFG) (Bresnan 2001)
Xerox Linguistic Environment (XLE) (Butt et al. 1999)

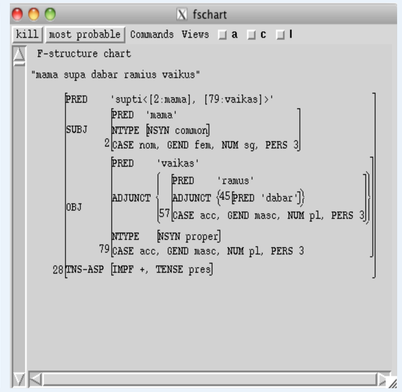
Number of rules: 9
- Phrase types: PP, NP, DP, AP, AdvP, IP, CP, vP, VP
- Declarative clauses
- Transpositions possible

Example:
mama supa dabar ramius vaikus
mom cradle now calm children
'Mom cradles the now calm children.'

C-Structure:



F-Structure:



Future Work
- Adding words into morphology
- Building generic roots → Universal morphological rules
- Expanding grammatic rules

More on the project: http://ltl.emich.edu/lithuanian/

Literature
Beesley, Kenneth R. und Lauri Karttunen (2003): *Finite State Morphology*. Stanford: CSLI Publications.
Butt, Miriam, Tracy Holloway King, Maria-Eugenia Nino, Frederique Segond (1999): *A Grammar Writers Cookbook*. Stanford: CSLI Publications.
Bresnan, Joan (2001): *Lexical-Functional Syntax*. Oxford: Blackwell.
Wiedemann, Oskar (2009): *Handbuch der litauischen Sprache: Grammatik, Texte, Wörterbuch*. BiblioLife.
Ambrazas, Vytautas (1997): *Lithuanian Grammar: Lietuviu kalbos gramatika*. Vilnius: Baltos Lankos Publishing House.