

# Grammar checker for Icelandic

NLP final project presentation

Kristján Halldórsson & Skúli Arnlaugsson



# Overview

- Introduction & background
- Method
- Product
- Problems
- Future work

# Introduction & background

- Grammar checkers are an important tool.
- To our knowledge no Icelandic grammar checkers exist.
  - A prototype was presented as a B.Sc. project in 2009 by K. Guðjónsdóttir.
- Three different methods to implement a grammar checker (Naber, 2003)
  1. Syntactical method (tree structure)
  2. Statistical method (patterns from a large corpus)
  3. Rule-based method

# Method

- The rule-based method uses handwritten rules.
- Has several advantages / disadvantages:
  - Limited to the written rules (never exhaustive).
  - Easy to configure and add rules.
  - Can look at partial sentences.

# Product

1. Grýla (Málgrýlan), grammar checker.
  - BNF descriptions for errors.
  - Implemented rules.
2. Plugin for Microsoft Word 2010.

# Grýla

- Uses IceNLP to tag and parse input text.
- Standalone module developed in Java.
- Easily accessible for other uses.

# Grýla BNF

- Output is a formalized error list according to a BNF description.
- Either returns an 'ok' or a list of found errors:
  - Format of error: Position Identifier Rulenummer [*List of corrections*]
  - Examples:
    - 2 góði 1 [góða]
    - 7 stór 2 [stórir]

# Grýla implemented errors

- We implemented 7 rules mostly dealing with noun phrases.
- Rule list:
  - 1. Disagreement in case within a noun phrase
  - 2. Disagreement in number within a noun phrase
  - 3. Disagreement in gender within a noun phrase
  - 4. Disagreement in gender between the subject and the complement
  - 5. Disagreement in number between the subject and the complement
  - 6. Disagreement in case in a prepositional phrase
  - 7. Disagreement in number between subject and related word



# MS Word plugin

- Uses the grammar checker module:
  - Sends the text to be checked to the grammar checker
  - Receives an error list or 'ok'.
  - Presents the errors to the user if any are found, underlines problem words and informs the user what rules are broken.
- Developed in Visual Studio 2010.

# Problems

- IceNLP
  - Sentences containing grammatical errors are not always parsed correctly. (ex. “Þær eru góð vinkonur”, IceTaggerFacade will tag “góð” as plural)
- Office development
  - Classes for the spelling and grammar are complicated and not easily accessible.

# Demo

# Future work

1. Java integration to C# / run Grýla as a service.  
(<http://www.janetdev.org/Part4UsingJavainaCApplication/tabid/118/Default.aspx>)
2. Use Apertium dictionaries and LTtoolbox to get suggestions for correct word forms.  
(<http://xixona.dlsi.ua.es/~fran/icelandic/>)
3. Seperate rules from implementation.
4. Make language independent (requires complete overhaul 😊).

# Thank you

Any questions?