# Creating New Language and Voice Components for the Updated MaryTTS Text-to-Speech Synthesis Platform

Ingmar Steiner[1–3], Sébastien Le Maguer[1,2]

[1]*Multimodal Computing and Interaction, Saarland University*
[2]*Computational Linguistics & Phonetics, Saarland University*
[3]*DFKI Language Technology Lab, Saarbrücken*

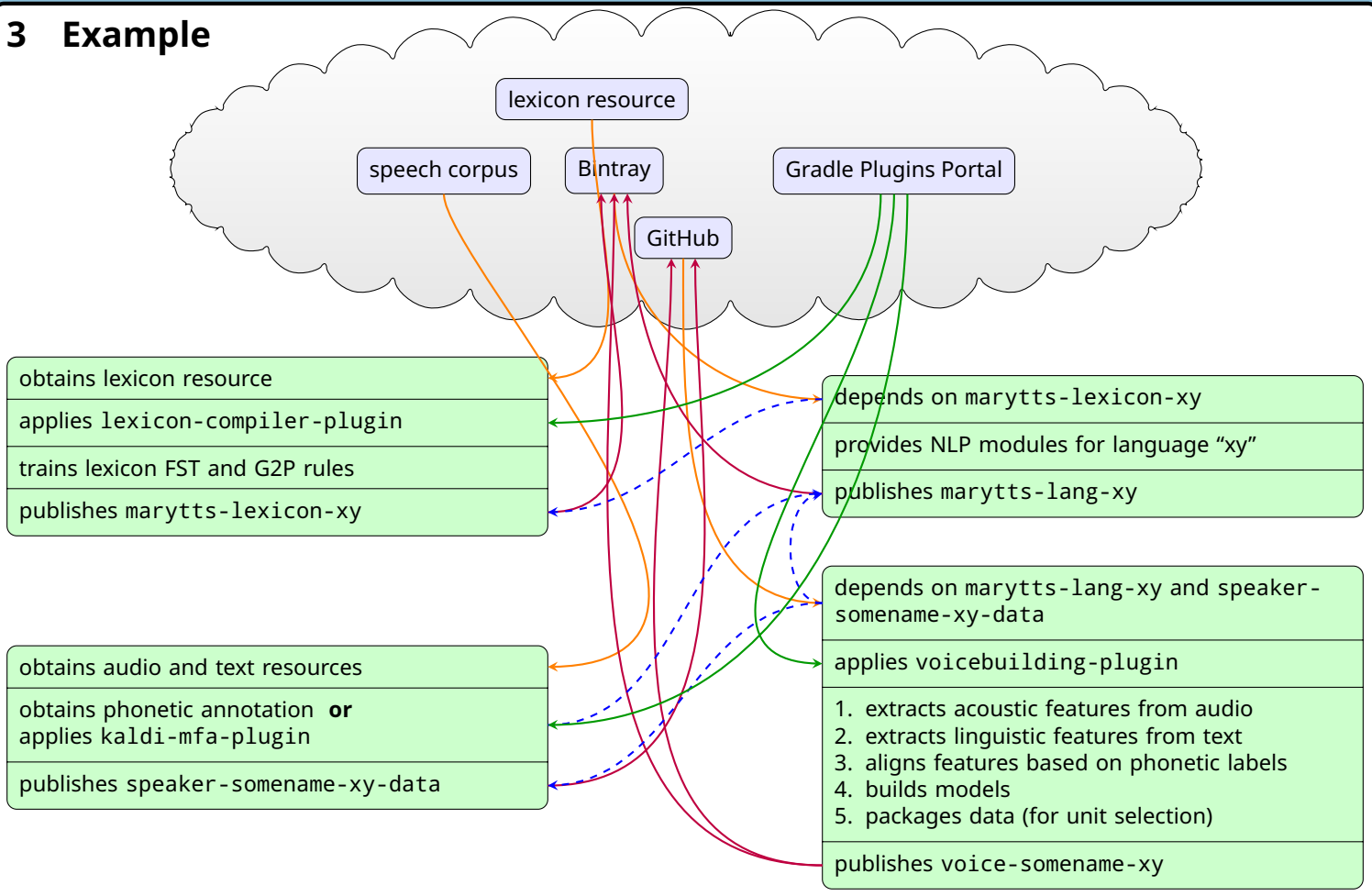
## 1 Introduction

- MaryTTS is a widely used, open-source text-to-speech synthesis (TTS) system
- A redesigned architecture has been implemented
- New component building process (languages and voices)
- Processing leverages build automation (Gradle)

## 2 Plugin-based voice building (VB)

1. Data preparation/forced alignment (Kaldi)
2. Acoustic feature extraction (Praat/SPTK/EST)
3. Linguistic feature extraction (MaryTTS)
4. Model building
   - Unit selection (acoustic models, data packaging)
   - Statistical parametric speech synthesis (HMM)

## 3 Example




## 4 New configuration mechanism

**default** given in the MaryTTS module
**voice** configured in voicebuilding project
**user** specified at runtime (takes precedence)

⇒ More flexible control over module pipeline, parameters, models assets, etc.

## 5 Conclusion

- New **language** and **voice** component building workflow
- Heavy use of state-of-the-art build automation
  + more efficient
  + more extensible (plugins)
- Entire process **redesigned and extended** (cloud-based)

UNIVERSITÄT DES SAARLANDES
Computational Linguistics & Phonetics
M2CI CLUSTER OF EXCELLENCE
DFKI
IDeaL SFB 1102