Integrating Grammar-Based Language Models into Domain-Specific ASR Systems

UNIVERSITÉ DE GENÈVE
FACULTÉ DE TRADUCTION ET D'INTERPRÉTATION

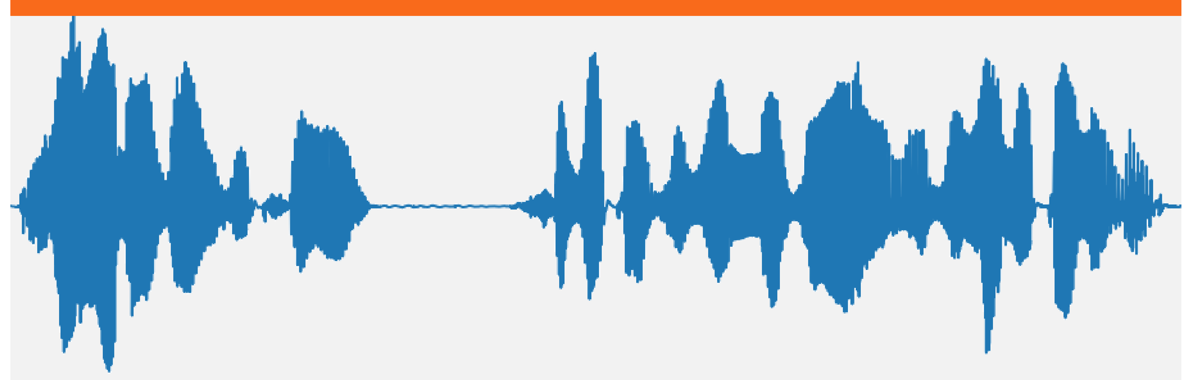
Advanced Language Processing School - ALPS

Lucía Ormaechea Grijalba

Ph.D. Candidate, *Multilingual Information Processing*

Department of Translation Technology – University of Geneva – Switzerland

Introduction and motivation



- **Language Models (LMs)** represent a vital component in the design of **ASR technologies**.
- **Domain-specific** ASR applications often require:

An efficient control of the outputs generated by the ASR system

- One possible solution is to rely on:

Probabilistic approaches ⇒
- ✓ **Generalizability**
- ✕ **Short memory span**
- ✕ **High-quality data not always available**

- Meaning that **more efficient modeling techniques** are needed.

Proposed solution

Implementing grammar-based LMs

✓ Large scope of constraints between words	✓ Parsing efficiency	✕ Lack of software to feed them into ASR toolkits

</> Development of **kaldi-grammar-compiler**

An OpenFST-based extension that transforms:

Regulus Lite (RL) grammars
- Finite-state-based.
- User-friendly syntax.

⇒ **Binary Finite State Transducers (FSTs)**
Readable LM format by Kaldi.

Kaldi – Speech Processing Toolkit
- Finite state framework for training and decoding algorithms.
- Highly usable and modifiable.

Methodology

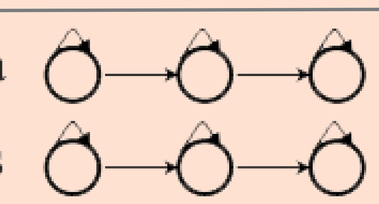
I. System overview

KEY POINTS	BENEFITS
Online (real-time) decoding	Better transcription accuracy compared to offline recognizers
Chain-based HMM-DNN models	Faster training and decoding

II. ASR architecture

Speech signal

Acoustic modeling (HC)



- Training set: **755 hours of French speech**, based on **Common Voice Dataset 7.0**.
- **Time Delay Neural Network** architecture (TDNN).
- MFCC features + speaker adaptation.

Lexicon modeling (L)

Grapheme-to-phone mapping based on:
- BDLEX French phonetical database.
- LIA_PHON automatic phonetizer.

WORD	PRON
vous	v u
avez	a v e
mal	m a l

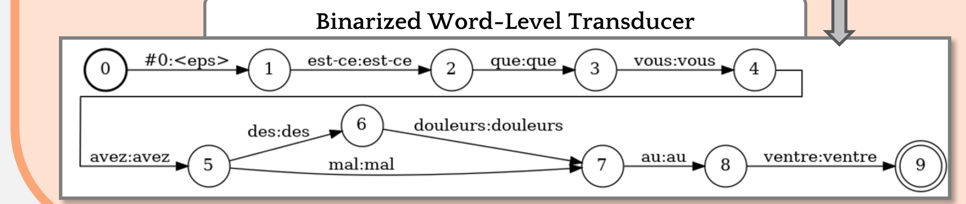
Grammar-based language modeling (G)

RL Fixed grammars

```
Utterance
Source est-ce que vous
avez ( des douleurs |
mal ) au ventre ?
EndUtterance
```

⇒ kaldi-grammar-compiler

Binarized Word-Level Transducer



Speech-to-text decoder (HCLG)

Est-ce que vous avez mal au ventre ?

Evaluation and results

I. Dedicated corpora for evaluation

	Description	Language	Speakers	Gender	Duration	Utterances	Words
MeDiCo – *Medical Discourse Corpus*	• Artificial spoken dataset. • Obtained via data collection campaign. • Including utterances related to **medical consultations**.	FR	14	9F, 5M	0h 41mn	713	≲6k
HomeAutomation	• **Voice command-based** spoken corpus. • Collected in realistic smart-home conditions.	FR	23	9F, 14M	1h 38mn	3114	≲10k

Results/Corpus	MeDiCo	HomeAutomation
Recognized words	5292 / 5598	9058 / 9639
Insertions	58	87
Deletions	34	295
Substitutions	214	199
WER (%)	5.47	6.03

II. Speech recognition results

- **Two different Kaldi ASR engines** were built.
- Both integrated a **fixed grammar** as **LM** in their **decoding graph (HCLG)**.
- Transcription accuracy was calculated in terms of **Word Error Rate (WER)**.

Summary and conclusions

Key points:

- **Introduced** an **extension** for **easily integrating fixed regular grammars** as **LMs** into **Kaldi**.
- Achieved **satisfactory results** in the experiments performed:
  - Both **MeDiCo** and **HomeAutomation** return a **significantly low WER** (5.47% and 6,03%, respectively).

This shows that **grammar-based ASR systems** obtain a **competitive** performance when applied in **constrained domain-specific applications**.

Further work

Some questions still need to be explored:

- Explore how to **leverage grammar knowledge**, so as to **specialize a neural-based LM**.
- **Generalize the input grammar format**, so as to:
  - Extend the applicability of our designed tool beyond the Regulus Lite syntax.

Acknowledgements

The authors would like to express their gratitude to those who kindly accepted to take part in the spoken data collection campaign.

References

- **Horndasch, A. *et al.*** (2016): How to Add Word Classes to the Kaldi Speech Recognition Toolkit. In *Text, Speech, and Dialogue.*
- **Povey, D. *et al.*** (2011): The Kaldi Speech Recognition Toolkit. *IEEE Workshop on Automatic Speech Recognition and Understanding.*
- **Rayner, M. *et al.*** (2018): Handling Ellipsis in a Spoken Medical Phraselator. In *Statistical Language and Speech Processing.*

This QR code links to the GitHub repository where we made available the **kaldi-grammar-compiler** extension. Some working examples are also provided.

Contact me: Lucia.OrmaecheaGrijalba@unige.ch