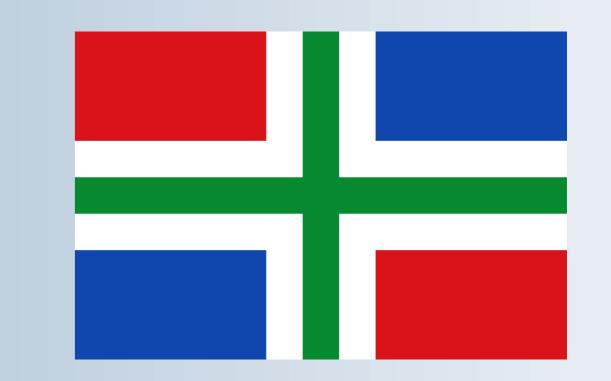
# **Automatic Machine Translation Dutch \leftrightarrow Gronings**



Rick Kosse Johan Bos Antonio Toral Ack. data providers: Goffe Jensma and Martijn Wieling





## Challenges

- Very little parallel data
- No spelling standards for Gronings
- Regional differences
- First MT system for NL→GR

### Method

- Three models: Word-, Character-Based, Bytepair encoding
- Corpora: Goud Volk and Martha
- ullet Goud Volk (8000 sentences) o Training, Dev, Test (In-domain)
- Martha (1000 sentences)  $\rightarrow$  Test (Out-domain)
- Trained with NMT-Keras (sequence-to-sequence)

#### **Gronings Characteristics**

- Word order auxiliary verbs in subordinated sentences
  - GR: Zeg mor davve nai' kommen willen.
  - NL: Zeg maar dat wij niet willen komen.
- Pro-drop questions second-person subject pronoun
  - GR: Hes dat doan?
  - NL: Heb je dat gedaan?
- Irregular verbs When the stem ends in an -l or an -r, only one -n is left in plural form.

GR: Not fiedelen but fiedeln

NL: Vioolspelen

### Parallel Corpora

Domain	Dutch
Domain	Dutcr

In-domain proces-verbaal moest er van opgemaakt worden.

jij hebt werkelijk een brutale mond. In-domain

't is al donker wanneer ze bij huis komen. In-domain

Out-domain hadden ze het over mijn verhaal?

Out-domain ze zei: nou, wees maar niet verdrietig hoor.

Out-domain heb jij ook eventjes gevaren.

#### Gronings

precesverbal mos d'r van opmakt wodden.

doe hes ja 'n brutoale mond.

't is aal duuster as ze op 't haim kommen.

haren ze t nou over mien verhoal?

nou, heb mor gain verdrait, zee ze. hest doe ook even boot jevoaren.

## **Evaluation** (BLEU)

System	In-domain	<b>Out-domain</b>
Baseline	18.09	10.48
Word	18.10	4.47
Char	23.68	10.76
BPE-50	23.32	9.28
BPE-100	25.12	8.83
BPE-250	24.55	5.10
BPE-500	23.43	3.31
BPE-1000	21.58	1.05

### **Translation Results**

- procht-verboal mos der van opgemoakt wodden.
- doe hest waarkelek 'n brutale mond.
- 't is aal donker ze bie hoes komen.
- hadden ze 't over mien verhoal?
- ze zee: nou, wees moar nait verdrietig heur.
- hes doe ook eventjes gevaren.

#### **Future Work**

- Create monolingual silver data ("backtranslate")
- Add silver data to training data and re-train
- ullet Apply same technique for Gronings o Dutch



