



MaChAmp: Multi-task Learning to the Rescue in Resource Scarce Scenarios



THE WALL STREET JOURNAL

© 1981 Dow Jones & Company, Inc. All Rights Reserved.

VOL. CXCVII NO. 14 ★ ★

STANDARD EDITION

WEDNESDAY, JANUARY 21, 1981

PROVIDENCE, NEW JERSEY

35 CENTS

After the Crisis

Torn U.S.-Iranian Ties
Won't Heal for a While
Despite Hostage Pact

Mutual Recriminations Seen
Likely, but in Long Run
'Normal' Relations Seen

'Stay Apart, Let Time Pass'

By KENNETH BLOOMBERG
WASHINGTON—The long hostage ordeal that has so poisoned U.S.-Iranian relations is over, but its bitter residue will block American foreign-policy goals in Iran for the indefinite future.

After nearly 15 months of captivity, the 52 American hostages held Tehran yesterday for Algeria. In return, the U.S. released about \$1.1 billion in Iran's frozen assets in U.S. banks and their European branches by President Carter on Nov. 19, 1979.

That swap could eventually open the way for a more effective American policy to

For details of the hostages' release, see story on page 2.

check Soviet influence in Iran, long a buffer between the Soviet Union and the Persian Gulf oil states. Meanwhile, President Reagan assumed office yesterday without the precedent-setting of the hostage issue, which has shackled American foreign policy, especially in the Middle East, for over a year.

What's News

Business and Finance

THREE MAJOR BANKS reported mixed results for the fourth quarter. Citicorp operating net plunged 30% to \$96 million despite a \$37 million after-tax gain from the sale of real estate and lease residuals. But Manufacturers Hanover had a 10% increase to \$58.5 million, and Wells Fargo's profit rose 7% to \$34.4 million.

First Pennsylvania Corp. had a fourth quarter loss of \$68.8 million, due mostly to a \$10.5 million expansion of loan loss provisions and \$11.5 million to settle holder suits. The bank holding company's loss for the year totaled \$184.1 million.

American Telephone and the government may both see a victory in the "taxable income" reached to settle their antitrust case. In return for divesting itself of part of Western Electric and some operating units, AT&T may be allowed to enter the data processing field.

Cincinnati Bell canceled a \$40 million note offering only a day before the securities were to be delivered to investors. The surprise move came after reports that a pending settlement

World-Wide

THE HOSTAGES FLEW to Algeria as President Reagan halted their freedom.

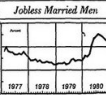
Two Algerian jets left Tehran at 10:35 p.m. EST and refueled in Athens, carrying the former captives and Algerian diplomats. In Algiers, the 52 Americans were welcomed by U.S. legislators, and the group chafed in an airport hangar. A brief envoy said the hostages were jubilant leaving Iran, where conditions demeaned the U.S.

The release came after the Carter administration closed up a late dispute over an appendix to the U.S.-Iranian accord. The U.S. agreed to waive \$1 billion of frozen Iranian assets to the bank of England for transfer to Tehran. Washington's sanctions against Iran are to be lifted as well.

Antes said Carter expressed that Iran delayed the hostages' departure until after the change of Presidents in case his own more embarrasingment.

Families cheered and wept at word of their captive relatives' return. Bells at the city hall in Albany, N.Y., rang 30 times at 10:35 p.m. in Albany, N.Y., civil defense sirens blared for 44 seconds, one second for each day the Americans were held prisoner.

CARTER PREPARED to greet five captives at a hospital in West Germany. Having passed along the precaution, Carter briefly returned home to Plains, Ga., to doubt if at any time in our history more prayers have reached heaven. He said Carter will be Reagan's emissary while three Americans recuperate at the U.S. mili-



UNEMPLOYMENT among married men fell to 4.3% of the labor force in December from a revised 4.8% the preceding month, the Labor Department reports.

Live, From St. Paul, Here's 'A Prairie Home Companion'

A Lazy Two Hours of Music And Small-Town Humor Scores Big on Public Radio

By LAWRENCE DINKOWSKI
and reporter of The Wall Street Journal
LAKE WOBEGON, Min.—If you've ever heard of this town, you know that it is the home of Ralph's Pasty Good Grocery, Our Lady of Perpetual Responsibility Church and the Shrine of the Unknown Homeless.

And that it can't be found on any map but only on your radio for two hours every Sunday evening.

The mythical town of Lake Wobegon (pronounced wuh-bog-on) comes to life each week on "A Prairie Home Companion," a two-hour radio show with a tiny Midwest-

Tax Report

A Special Summary and Forecast of Federal and State Tax Developments

FOREIGN-CONVENTION rules are eased—with some exceptions.

Congress attached various amendments in 1976 by limiting user's deductible business conventions abroad to two a year and imposing spending and reporting requirements. A new law repeals the two-meeting limit, eases other restrictions and allows any foreign meeting that's "as reasonable" to have outside North America as well. However, unions Richard Humber at Price, Waterhouse & Co., that means a meeting deductible under the old two-year rule may not be under the new requirements.

That test removes the previous restrictions of the organization and the meeting, the residence of the members, and the site of their other gatherings. Meetings on cruise ships aren't deductible at all. The law broadens the deductible North American area to include Canada and Mexico, besides the U.S. and its possessions; that should stimulate meetings there.

The pending U.S.-Canada tax treaty would give more of the same benefits. Missouri rules Canada may have made earlier recoussions it never's face.

THEY'LL KEEP THE HOUSE and all the pills, too.

The pills from setting your home normally is taxable, unless you put it into a new home by a certain time. But up to \$100,000 of gain on selling your principal residence will be excluded once from your gross income if you've over 55 and have owned your home there for three of the five prior years. See page 16, and his wife.

Fire Hazard

Safety Officials Fear Skyscraper Holocaust Could Kill Thousands

They Cite Buildings' Design And Location, Lax Codes, Poison Gas From Plastic

Owners Note Record Is Good

By FRANK R. ROSTON
Special Reporter for The Wall Street Journal
Fire fighters say it moves and over. In some they can get to know: A major fire disaster is likely any day, a disaster that will rival an earthquake or flood in death and destruction.

Their nightmare is a fire that rages wild through a skyscraper during working hours, trapping thousands of people on upper floors. Hundreds, even thousands could be asphyxiated or burned to death in a few minutes.

Some think it's more than likely. "I'm not talking about probability. I'm talking about certainty," says Michael F. Laffey, a former captain and 7½-year veteran of the New York City Fire Department. He is currently safety officer at Citicorp's 59 story Manhattan office tower.

Why the Concern?

Speculators like Mr. Laffey can tick off their concerns:

- More people than ever work in office space high-rise office buildings are going up, many in suburbs likely to be served by fire

What to do about non-standard (or *non-canonical*) language in NLP

Barbara Plank
University of Groningen
b.plank@rug.nl

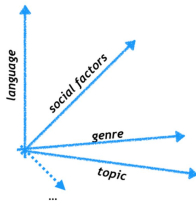
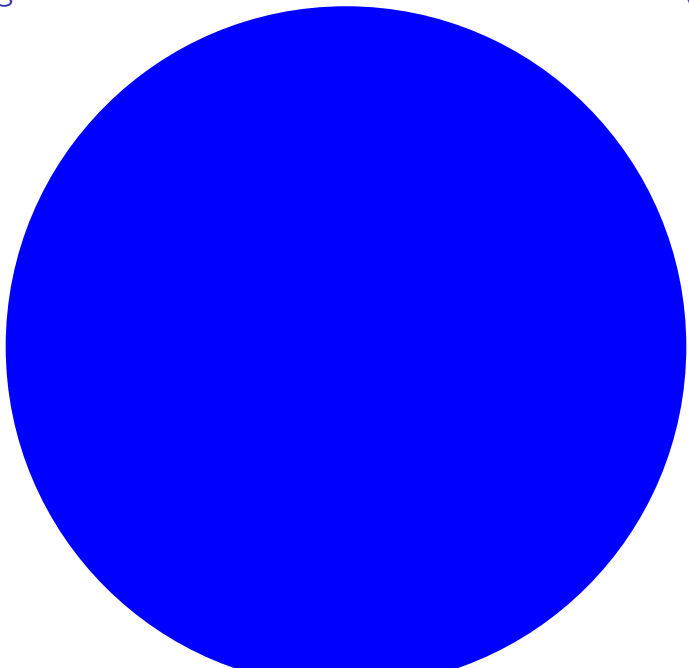


Figure 2: What's in a *domain*? Domain is an overloaded term. I propose to use the term *variety*. A dataset is a sample from the *variety space*, a unknown high-dimensional space, whose dimensions contain (fuzzy) aspects such as language (or dialect), topic or genre, and social factors (age, gender, personality, etc.), amongst others. A domain forms a region in this space, with some members more prototypical than others.

Language varieties that are annotated for NLP tasks(in red)



What can we do?


- ▶ Annotate more?
 - ▶ Language is not static!
- ▶ Cross-domain, cross-lingual learning

Multi-task learning to the rescue!

Standard in NLP:

- ▶ Pre-train a **language model** on raw data (billions of words)
- ▶ Fine-tune the language model on NLP-annotated data (thousands of words)

What is a language model?

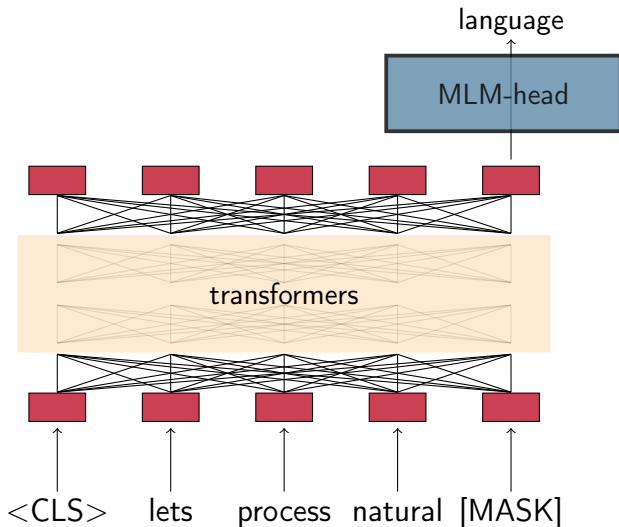
i  ved

Øvelse 34. Kig på stuen side 76 og indsæt passende præpositioner.

1. Der står nogle bøger _____ reolen.
2. Der står et fjernsyn _____ sofaen.
3. Der ligger et gulvtæppe _____ sofabordet.
4. Der står en kop _____ telefonen.
5. Der står en vase _____ sofabordet.
6. Der hænger nogle billeder _____ sofaen.

vase

What is a language model?



Framework: MaChAMp

Massive Choice, Ample Tasks (MACHAMP):



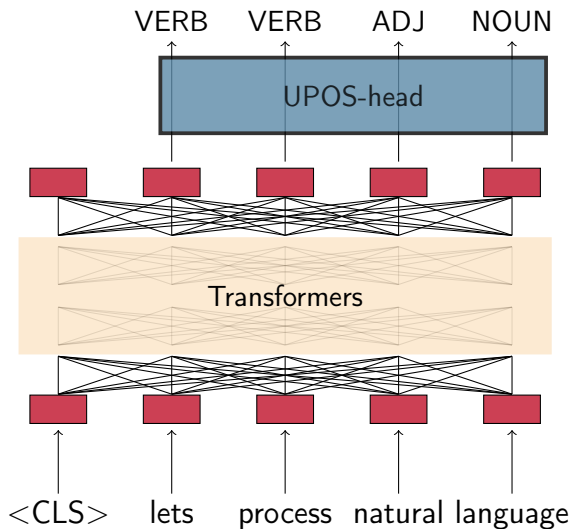
A Toolkit for Multi-task Learning in NLP



Rob van der Goot 🇳🇱 **Ahmet Üstün** 🇳🇱 **Alan Ramponi** 🇮🇹🇮🇹 **Ibrahim Sharaf** 🇪🇬
Barbara Plank 🇩🇪


IT University of Copenhagen 🇩🇰 University of Groningen 🇳🇱 University of Trento 🇮🇹
Fondazione the Microsoft Research - University of Trento COSBI 🇮🇹 Factmata 🇮🇹
robv@itu.dk, a.ustun@rug.nl, alan.ramponi@unitn.it
ibrahim.sharaf@factmata.com, bapl@itu.dk

MaChAmp



- ▶ This is the default setup for all NLP tasks these days; sharing happens over time: MLM \Rightarrow TGT task
- ▶ MaChAmp can do much more!, we mainly focus on multi-task learning after the first pre-training

MaChAmp #001



Abilities:

joint training	loss weighing
sequential training	layer attention
dataset embeddings	dataset smoothing

MaChAmp is a multi-task NLP toolkit, it can seemingly effortlessly handle multiple NLP tasks simultaneously. It supports a wide variety of NLP tasks, and can easily handle multiple datasets at once.

More information on:
machamp-nlp.github.io

IT UNIVERSITY OF CPH

Examples of tasks

Input	Output
classification	
Smell ya later!	negative
mlm	
Gotta [MASK] em all	catch
multiclas	
That will be 5\$	inform request
multiseq	
I never caught Snorlax	per:1 n:sin _ tens:past n:sin
regression	
You're playing cats	1.2
seq	
I want to be the best	PRN VB PART AUX DT ADJ
seq_bio	
Ash from Pallet Town	Ash:PERS Pallet Town:LOC
tok	
Gary, Gary, he's the man.	Gary , Gary , he 's the man .
dependency	
Brock wants to fight	Brock wants to fight

Rob van der Goot

Examples of tasks

Input

Output

classification

Smell ya later!

negative

mlm

Gotta [MASK] em all

catch

multiclas

That will be 5\$

inform|request

multiseq

I never caught Snorlax

per:1|n:sin _ tens:past n:sin

regression

You're playing cats

1.2

seq

I want to be the best

PRN VB PART AUX DT ADJ

seq_bio

Ash from Pallet Town

Ash:PERS Pallet Town:LOC

tok

Gary, Gary, he's the man.

Gary , Gary , he 's the man .

dependency

Brock wants to fight

Brock wants to fight

```
graph TD
    root((root)) --- xcomp[xcomp]
    root --- nsubj[nsubj]
    root --- mark((mark))
    xcomp --- mark
```

The diagram shows a dependency parse tree for the sentence "Brock wants to fight". The root node is labeled "root". It has three children: "nsubj", "xcomp", and "mark". The "xcomp" node has a child "mark". The "mark" node is connected to the "nsubj" node by a curved arrow pointing from "mark" to "nsubj".

xSID: Cross-lingual Slot and Intent Detection

Rob van der Goot, Ibrahim Sharaf, Aizhan Imankulova, Ahmet Üstün, Marija Stepanović, Alan Ramponi, Siti Oryza Khairunnisa, Mamoru Komachi and Barbara Plank



Example: Languages in EU covered by voice assistants



*as of March, 2020



87

* Slide by Barbara Plank

Slot and Intent Detection

I'd like to see the showtimes for Silly Movie 2.0 at the movie house

Intent: SearchScreeningEvent

xSID

ar أود أن أرى مواعيد عرض فيلم Silly Movie 2.0 في دار السينما
da Jeg vil gerne se spilletiderne for Silly Movie 2.0 i biografen
de Ich würde gerne den Vorstellungsbeginn für Silly Movie 2.0 im Kino sehen
de-st I mecht es Programm fir Silly Movie 2.0 in Film Haus sechn
en I'd like to see the showtimes for Silly Movie 2.0 at the movie house
id Saya ingin melihat jam tayang untuk Silly Movie 2.0 di gedung bioskop
it Mi piacerebbe vedere gli orari degli spettacoli per Silly Movie 2.0 al cinema
ja 映画館の Silly Movie 2.0 の上映時間を見せて。
kk Мен Silly Movie 2.0 бағдарламасының кинотеатрда көрсетілім уақытын көргім келеді
nl Ik wil graag de speeltijden van Silly Movie 2.0 in het filmhuis zien
sr Želela bih da vidim raspored prikazivanja za Silly Movie 2.0 u bioskopu
tr Silly Movie 2.0'in sinema salonundaki seanslarını görmek istiyorum
zh 我想看 Silly Movie 2.0 在影院的放映

xSID

ar أود أن أرى مواعيد عرض فيلم Silly Movie 2.0 في دار السينما
da Jeg vil gerne se spilletiderne for Silly Movie 2.0 i biografen
de Ich würde gerne den Vorstellungsbeginn für Silly Movie 2.0 im Kino sehen
de-st I mecht es Programm fir Silly Movie 2.0 in Film Haus sechn
en I'd like to see the showtimes for Silly Movie 2.0 at the movie house
id Saya ingin melihat jam tayang untuk Silly Movie 2.0 di gedung bioskop
it Mi piacerebbe vedere gli orari degli spettacoli per Silly Movie 2.0 al cinema
ja 映画館の Silly Movie 2.0 の上映時間を見せて。
kk Мен Silly Movie 2.0 бағдарламасының кинотеатрда көрсетілім уақытын көргім келеді
nl Ik wil graag de speeltijden van Silly Movie 2.0 in het filmhuis zien
sr Želela bih da vidim raspored prikazivanja za Silly Movie 2.0 u bioskopu
tr Silly Movie 2.0'in sinema salonundaki seanslarını görmek istiyorum
zh 我想看 Silly Movie 2.0 在影院的放映

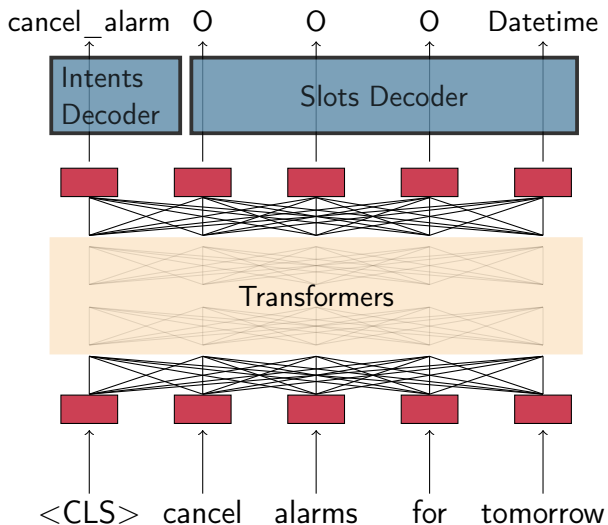
- ▶ Lithuanian
- ▶ Bavarian
- ▶ Swiss German
- ▶ Neapolitan

Experiments

Baselines

- ▶ Baseline: language model with joint intent+slots

Baseline



Experiments

Baselines

- ▶ Baseline: language model with joint intent+slots
- ▶ Stronger baseline: translate training data to target language (NMT-TRANSFER)

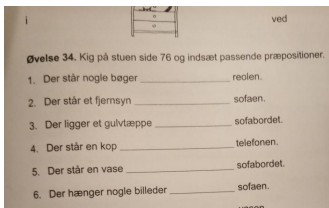
Experiments

Baselines

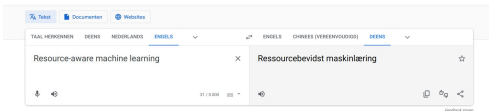
- ▶ Baseline: language model with joint intent+slots
- ▶ Stronger baseline: translate training data to target language (NMT-TRANSFER)

New models:

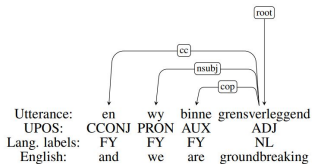
- ▶ Train on auxiliary task in target language:
 - ▶ Masked language modeling (AUX-MLM)
 - ▶ Neural machine translation (AUX-NMT)
 - ▶ UD-parsing (AUX-UD)



► MLM:



► NMT:



► UD-parsing:

Experiments

Evaluate 2 language models

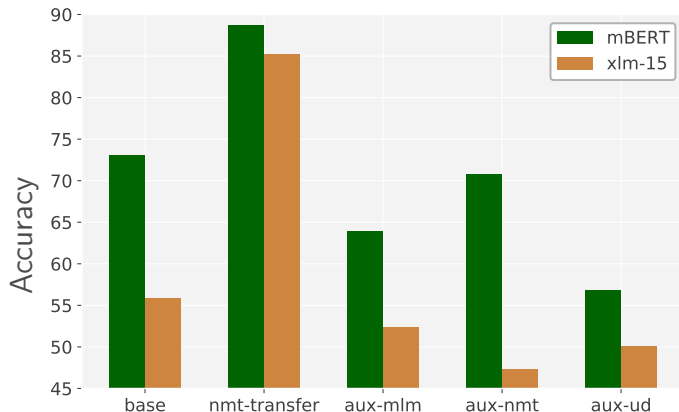
- ▶ mBERT: trained on 104 languages (12/13)
- ▶ XLM15: trained on 15 languages (5/13)

Results

model	Time (minutes)
base	46
nmt-transfer	5,213
aux-mlm	193
aux-nmt	373
aux-ud	79

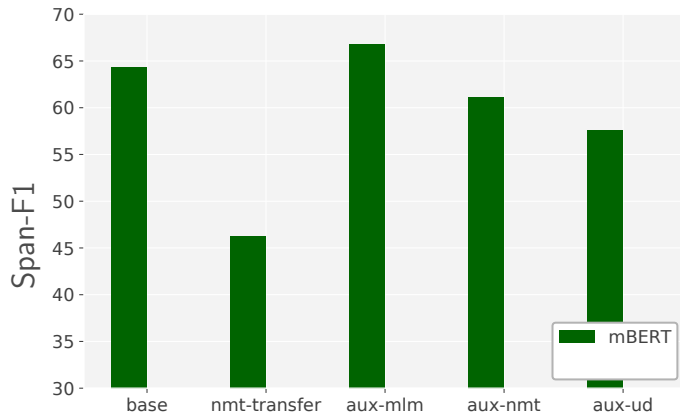
Table: Average minutes to train a model on a V100, averaged over all languages and both embeddings. For nmt-transfer we include the training of the NMT model.

Results (intents)



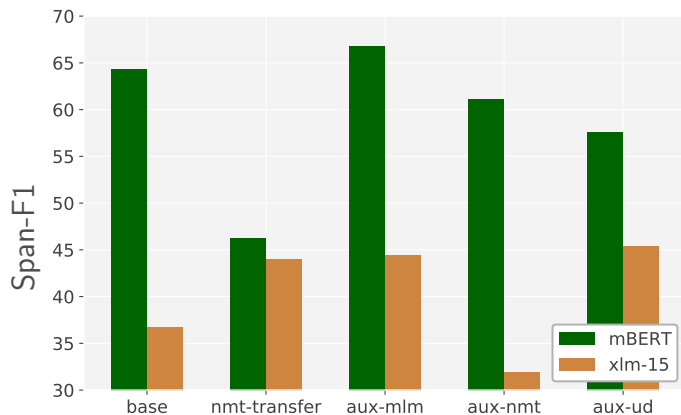
- ▶ NMT-transfer is hard to outperform, but costly
- ▶ Even baseline hard to beat

Results (slots)



- ▶ NMT-transfer performs bad (due to alignment)
- ▶ Seen languages: only MLM helps

Results (slots)



- ▶ NMT-transfer performs bad (due to alignment)
- ▶ Seen languages: only MLM helps
- ▶ Unseen languages: More explicit tasks (UD) are faster and lead to better performance

How to implement this

Requirements:

- ▶ Training data
- ▶ In tab-separated format
- ▶ Couple hours of GPU time

How to implement this

```
# text-en = Is it cloudy today?  
# text = Er det overskyet i dag?  
# intent = weather/find  
1      Er      0  
2      det     0  
3      overskyet      B-weather/attribute  
4      i        B-datetime  
5      dag      I-datetime  
6      ?        0
```

How to implement this

```
xsid.json:
{ "xSID": {
  "train_data_path": "xsid-0.5/en.train.conll",
  "validation_data_path": "xsid-0.5/en.dev.conll",
  "word_idx": 1,
  "tasks": {
    "slots": {
      "task_type": "seq_bio",
      "column_idx": 2
    }
    "intents": {
      "task_type": "classification",
      "column_idx": -1
    }
  }
}
}
```

```
python3 train.py --dataset_configs xsid.json
```


How to implement this

```
mlm-da.json:
{ "MLM": {
  "train_data_path": "data/spoken/da.train.txt",
  "validation_data_path": "data/spoken/da.dev.txt",
  "tasks": {
    "language_modeling": {
      "task_type": "mlm",
    }
  }
}
}
python3 train.py --dataset_configs xsid.json mlm-da.json
```

How to implement this

```
mlm-da.json:
{ "MLM": {
  "train_data_path": "data/spoken/da.train.txt",
  "validation_data_path": "data/spoken/da.dev.txt",
  "tasks": {
    "language_modeling": {
      "task_type": "mlm",
    }
  }
}
}
```

```
python3 train.py --dataset_configs xsid.json mlm-da.json
```

Need better performance?: try other language models.

Interested in SOTA NLP?:

- ▶ Try MaChAmp:
<https://github.com/machamp-nlp/machamp.git>
- ▶ Get in touch (slack/e-mail)

