

Moses

M. Fabianová, A. Štromajerová, M. Vaněk

Osnova

1. Trocha historie
2. Co je to Moses?
3. V čem je Moses jiný/lepší než ostatní SMT překladače?
4. Využití, příklady, srovnání

Trocha historie

50. léta	první systémy založené na pravidlech
1966	ALPAC => snížení grantů na výzkum SP
80. léta	první komerční systémy
90. léta	frázový překlad
2000 +	prudký vzestup strojového překladu
2006	představení projektu Moses

Co je to Moses?

Moses

= nejstarší a nejrozšířenější open-source systém pro statistický strojový překlad

- umožňuje automatickou přípravu překladových modelů pro jakoukoliv dvojici jazyků
- flexibilní

Co je potřeba?

- sbírka překladů (paralelní korpus)
==> efektivní vyhledávací algoritmus najde nejpravděpodobnější překlad požadovaného textu mezi exponenciálním počtem dostupných možností



MOSES

statistical
machine translation
system

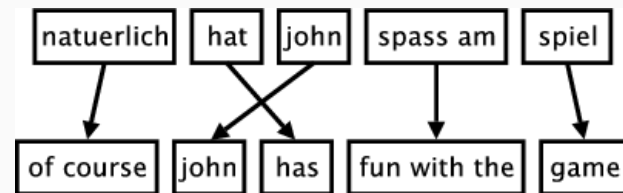
Rozšíření

- Vývoj: **University of Edinburgh** (+ Itálie, ČR (UK v Praze),...)
- 7000 návštěv webových stránek/měsíc
- a. akademické využití
 - strojový překlad
 - výzkum
 - vyhodnocování kampaní
- b. komerční využití
 - úprava podle požadavků uživatele a následný prodej
 - společnosti zajišťující jazykové služby

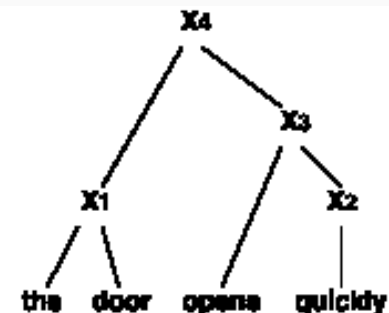
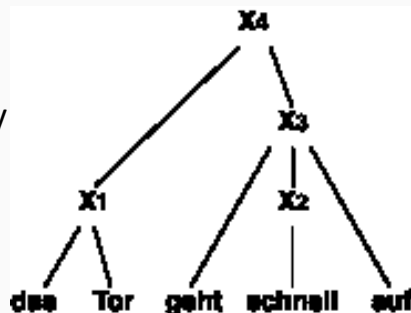


Moses umožňuje

- dva typy překladů:
 - a) TREE-BASED = model založený na syntaktických stromech
 - b) PHRASE-BASED (frázový překlad) = překlad celých sekvencí slov



- možnost budovat vlastní překladové systémy na základě systému Moses
- plná kontrola nad procesem překladu



Online demo

<http://demo.statmt.org/>

Poznátky:

- překlad chvíli trvá
- poměrně nepřesné (čeština)
- mělo by fungovat lépe pro: angličtina, španělština, němčina, čínština

V čem je Moses jiný/lepší než ostatní SMT překladače?

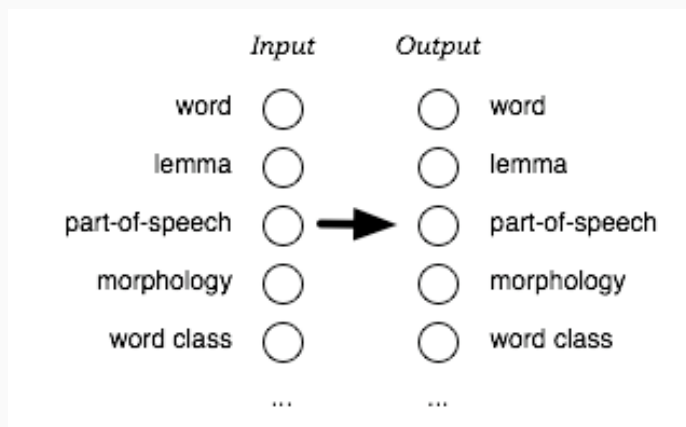
Open source/Free software

- většina předchozích systémů SMT proprietární/s restriktivními licencemi
- proprietární software
 - nespravedlivý
 - „černá skříňka“ (black box)
 - neprospěšný celé komunitě
 - nemožné jej dále zkoumat, rozvíjet, vylepšovat
- svobodný software
 - dostupný pro všechny
 - bez nutnosti vymýšlet něco, co již bylo vymyšleno
- Moses je svobodný software – distribuován pod licencí LGPL od GNU



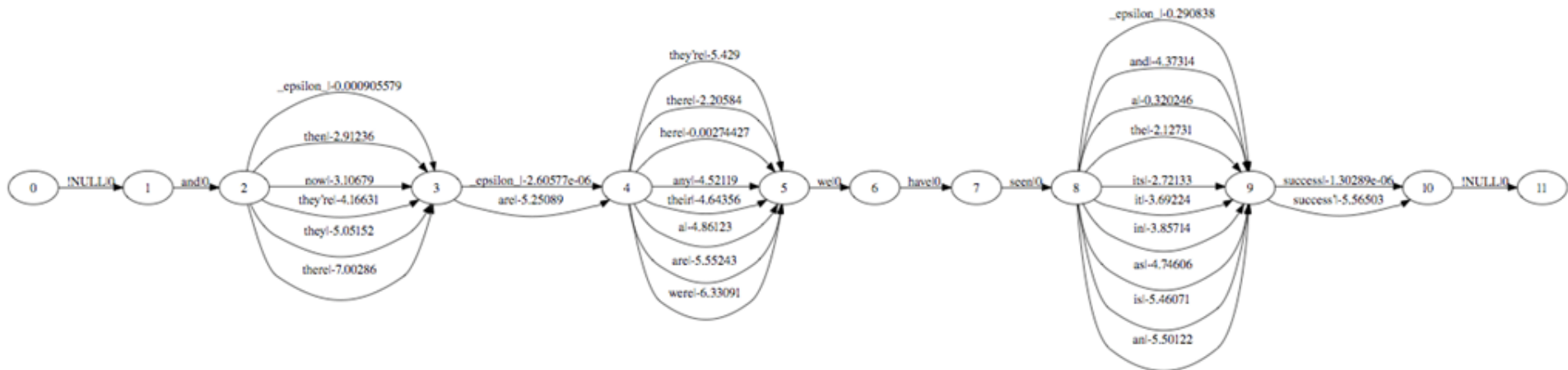
Factored translation

- integrování přídatných lingvistických informací do překladového modelu
 - morfologie, syntax, sémantika
- na úrovni slov vs. při pre- či post-processingu
- mnoho aspektů překladů záleží na těchto lingvistických informacích – pořadí slov ve větě (syntax), shoda přídatného jména s podstatným jménem (morfologie), apod.
- slovo-token → slovo-vektor faktorů



Confusion network decoding

- confusion networks - vážené orientované grafy (slovo a jeho pravděpodobnost)
- vstup
 - SMT systémy obecně - jednoduchá sekvence slov, JEDNA vstupní hypotéza
 - Moses - confusion networks, VÍCE vstupních hypotéz
- použití např. při zpracování vstupů ze systémů rozpoznávání řeči, morfologické analýzy, apod.



Efektivní struktura dat

- systémy strojového překladu – velké nároky na hardware
- Moses – načítání dat překladové tabulky
 - načtení VŠECH dat do paměti – rychlý překlad, zabere hodně paměti
 - načtení ON DEMAND – pomalejší překlad, ale zabere velmi málo paměti
 - využití prefixových stromů (rychlejší, potřebují méně paměti)
- Moses - jazykové modely
 - nejčastěji používané SRILM
 - další: KenLM, IRSTLM, RandLM – menší paměťové nároky než SRILM

Využití, srovnání, příklady

Využití všeobecně

- cílem obvykle není vytvořit všeobjímající překladový systém
- zaměření zejména na specifické uživatele a domény
 - možnost natrénování na specifických datech a tím zvýšení úspěšnosti
- výsledky Moses slouží jako baseline nové systémy
- výzkum financován Evropskou unií => zaměření zejména na evropské jazyky

Komerční využití

- komerční využití Moses – TAUS
- hojně využíván jako základ pro reálné překladové systémy
- Lets MT, Moses for Mere Mortals – usnadnění vytvoření vlastního MT systému
- computer-aided translation
- malé paměťové nároky – v budoucnu pro mobilní zařízení?

Srovnání úspěšnosti Moses vs. Google Translate

- ve všeobecné doméně u častých jazykových párů zpravidla vítězí GT – výhoda velká databáze
- v omezené doméně vítězí Moses – možnost specifického natrénování
- Khalilov, Choudhury 2012:
 - čínština–angličtina: Moses lepší než GT (2x vyšší BLEU, GTM +7 %)
 - angličtina–čínština: Moses výrazně lepší než GT (4x vyšší BLEU, GTM +20 %)
- Moses-based systém Chimera (ÚFAL) slibuje vyšší úspěšnost než GT

Příklady překladu – kontext

Koupili jsme losy a soby.

Koupili jsme soby a losy.

Na počátku stvořil Bůh nebe a zemi.

Bible - Na počátku stvořil Bůh nebe a zemi.

Moses:

We bought tickets and reindeer.

We bought reindeer and elk.

At the beginning **of the God** created the heaven and earth.

The Bible - **in** the beginning **God** created the heaven and earth.

Srovnání překladu 1 – běžný text

Miloš Zeman (* 28. září 1944 Kolín) je český politik, ekonom, prognostik a třetí prezident České republiky. Úřadu se ujal složením slibu dne 8. března 2013.

Moses: Milos Zeman (* 28 September 1944 Cologne) is the czech politician, economist, expert and the third president of the Czech Republic. He took the oath of office on 8 March 2013.

Google Translate: Milos Zeman (* 28 September 1944 Cologne) is a Czech politician, economist and forecaster, and the third president of the Czech Republic. Office took the oath of office on 8 March 2013.

Bing Translator: Miloš Zeman (born September 28, 1944 in Kolín) is a Czech politician, Economist, writer and the third President of the Czech Republic. The Office took the oath the day 8. March 2013.

Srovnání překladu 2 – idiomy

Starý mládenec zjistil, že tu něco neklape, a tak z čista jasna vzal nohy na ramena.

Moses:

The old boy found that something was amiss, and then suddenly took on his shoulders.

Google Translate:

Bachelor realized that something was amiss, so out of the blue took to his heels.

Bing Translator:

A Bachelor has found that there's something wrong, and so out of the blue he took her legs on his shoulders.

Srovnání překladu 3 – homonyma

Ženu ženu holí.

Anička má svačinu a Honzík jí ji jí.

Moses:

The woman woman with a stick.

Annie has a snack and Honzík her her.

Google Translate:

Woman woman shaves.

Annie has a snack and Johnny had it to her.

Bing Translator:

Woman to woman.

Annie has a snack and Habila her her.

Srovnání překladu 4 – robustnost

Zapoměl jsem.

Je to výjimka.

Vyzkoušela bys ty šaty?

Moses:

I forgot.

It is an exception.

Bysis tried the dress?

Google Translate:

I forgot.

It is the exception.

Would you tried the dress?

Bing Translator:

I forgot.

It is the exception.

Tried to marry the dress?

KONEC

Děkujeme za pozornost.