Machine Translation

08: Morphology

Rico Sennrich
(slide credit: Adam Lopez)

University of Edinburgh

Example

Pair	Probability
P($Haus\|house$)	0.2110788
P($Hause\|house$)	0.0384082
P($Ordnung\|house$)	0.0251976
P($innerhalb\|house$)	0.0230789
P($bringen\|house$)	0.0227120
P($vor\|house$)	0.0205860
P($ihr\|house$)	0.0183205
P($gute\|house$)	0.0165660
P($Einen\|house$)	0.0164834
P($rehabilitieren\|house$)	0.016365
P($entgegnen\|house$)	0.016365
P($geboten\|house$)	0.0162445
P($gewöhnt\|house$)	0.0166643
P($27\|house$)	0.0161492
P($erweitern\|house$)	0.0151807
P($Dafür\|house$)	0.0151306
P($notwendig\|house$)	0.0142576
P($begegnen\|house$)	0.0138590
P($Arbeit\|house$)	0.01373755
P($sicheren\|house$)	0.0131492

Pair	Probability
P($Häuser\|houses$)	0.22393465
P($nach\|houses$)	0.02897585
P($Daches\|houses$)	0.02893843
P($halten\|houses$)	0.02756981
P($beihilfefähig\|houses$)	0.02536119
P($Wohnungen\|houses$)	0.02536087
P($Privathäuser\|houses$)	0.02513098
P($EFRE - Verordnung\|houses$)	0.02480666
P($immer\|houses$)	0.02395412
P($Palästinenserführer\|houses$)	0.02318363
P($Parlamentsreden\|houses$)	0.02318363
P($abzureißen\|houses$)	0.02318363
P($gleichmachen\|houses$)	0.02318363
P($Erdboden\|houses$)	0.02318363
P($bombardieren\|houses$)	0.02318363
P($genauen\|houses$)	0.0231831
P($schießen\|houses$)	0.0231829
P($anderem\|houses$)	0.02315655
P($fest\|houses$)	0.0222202
P($gezielt\|housesng$)	0.02200406

Example

Pair	Probability
P($Haus \mid house$)	0.2110788
P($Hause \mid house$)	0.0384082

Pair	Probability
P($Häuser \mid houses$)	0.22393465

Example

Pair	Probability
P($Haus\|house$)	0.2110788
P($Hause\|house$)	0.0384082

Pair	Probability
P($Häuser\|houses$)	0.22393465

… fanden sie sie auf den Stufen unseres Hauses sitzend …

Example

Pair	Probability
P($Haus\|house$)	0.2110788
P($Hause\|house$)	0.0384082

Pair	Probability
P($Häuser\|houses$)	0.22393465

… fanden sie sie auf den Stufen unseres Hauses sitzend …

… they found her sitting on the steps of our house …

Example

Pair	Probability
$P(Haus \mid house)$	0.2110788
$P(Hause \mid house)$	0.0384082

Pair	Probability
$P(Häuser \mid houses)$	0.22393465

… fanden sie sie auf den Stufen unseres Hauses sitzend …

… they found her sitting on the steps of our house …

	Singular	Plural
Nominativ	das Haus	die Häuser
Genitiv	des Hauses	der Häuser
Dativ	dem Haus dem Hause	den Häusern
Akkusativ	das Haus	die Häuser

Example

Pair	Probability
P($Haus \mid house$)	0.2110788
P($Hause \mid house$)	0.0384082

Pair	Probability
P($Häuser \mid houses$)	0.22393465

… fanden sie sie auf den Stufen unseres Hauses sitzend …

… they found her sitting on the steps of our house …

	Singular	Plural
Nominativ	das Haus	die Häuser
Genitiv	des Hauses	der Häuser
Dativ	dem Haus dem Hause	den Häusern
Akkusativ	das Haus	die Häuser

different *inflections* (sometimes called *declensions* for nouns)

Example

Pair	Probability
$P(Haus\|house)$	0.2110788
$P(Hause\|house)$	0.0384082

Pair	Probability
$P(Häuser\|houses)$	0.22393465

… fanden sie sie auf den Stufen unseres Hauses sitzend …

… they found her sitting on the steps of our house …

	Singular	Plural
Nominativ	das Haus	die Häuser
Genitiv	des Hauses	der Häuser
Dativ	dem Haus dem Hause	den Häusern
Akkusativ	das Haus	die Häuser

different *case*

Example

Pair	Probability
P($Haus \mid house$)	0.2110788
P($Hause \mid house$)	0.0384082

Pair	Probability
P($Häuser \mid houses$)	0.22393465

… fanden sie sie auf den Stufen unseres Hauses sitzend …

… they found her sitting on the steps of our house …

	Singular	Plural
Nominativ	das Haus	die Häuser
Genitiv	des Hauses	der Häuser
Dativ	dem Haus dem Hause	den Häusern
Akkusativ	das Haus	die Häuser

different *number*

Example

Pair	Probability
$P(Haus \mid house)$	0.2110788
$P(Hause \mid house)$	0.0384082

Pair	Probability
$P(Häuser \mid houses)$	0.22393465

… fanden sie sie auf den Stufen unseres Hauses sitzend …

… they found her sitting on the steps of our house …

agreement

	Singular	Plural
Nominativ	das Haus	die Häuser
Genitiv	des Hauses	der Häuser
Dativ	dem Haus dem Hause	den Häusern
Akkusativ	das Haus	die Häuser

Morphology is *productive*

	Singular	Plural
Nominativ	das Haus	die Häuser
Genitiv	des Hauses	der Häuser
Dativ	dem Haus dem Hause	den Häusern
Akkusativ	das Haus	die Häuser

	Singular	Plural
Nominativ	das Buch	die Bücher
Genitiv	des Buchs des Buches	der Bücher
Dativ	dem Buch dem Buche	den Büchern
Akkusativ	das Buch	die Bücher

	Singular	Plural
Nominativ	der Computer	die Computer
Genitiv	des Computers	der Computer
Dativ	dem Computer	den Computern
Akkusativ	den Computer	die Computer

das Schloss translates as *the castle*.

Can you translate the previously unseen *die Schlösser*?

Inflection interacts with phonology

	-er	-ir	-re
Infinitival form	manger	choisir	descendre
Gloss	'eat'	'choose'	'descend'
1sg	mang+e	chois+is	descend+s
2sg	mang+es	chois+is	descend+s
3sg	mang+e	chois+it	descend+
1pl	mang+eons	chois+issons	descend+ons
2pl	mang+ez	chois+issez	descend+ez
3pl	mang+ent	chois+issent	descend+ent

Inflection interacts with phonology

a. üzüldünüz
üz-ül-dü-nüz
sadden-PASS-PST-2PL
'You became sad.' [tur]

b. sevildiniz
sev-il-di-niz
like-PASS-PST-2PL
'You were liked.' [tur]

Turkish vowel harmony

simple morphology: English

- Case *(e.g. nom., dat.)*
- Number *(sg., pl.)*
- Person *(1st, 2nd, 3rd)*
- Tense *(past, present)*

	singular	plural
nominative	I	we
oblique	me	us
possessive determiner	my	our
possessive pronoun	mine	ours
reflexive	myself	ourselves

Tense	I	you	he, she, it	we	you	they
Present	arrive	arrive	arrives	arrive	arrive	arrive
Past	arrived	arrived	arrived	arrived	arrived	arrived

more complex: German

- Inflections of the English definite determiner:

more complex: German

- Inflections of the English definite determiner: *the*

more complex: German

- Inflections of the English definite determiner: *the*
- Inflections of the German definite determiner:

	Singular			Plural
	Maskulinum	Femininum	Neutrum	–
Nominativ	der	die	das	die
Genitiv	des	der	des	der
Dativ	dem	der	dem	den
Akkusativ	den	die	das	die

Noun classes in Luganda

- Class I contains mainly people, although some inanimate nouns can be found in this class: musajja 'man', kaawa 'coffee'
- Class II contains all sorts of nouns but most of the concrete nouns in Class II are long or cylindrical. Most trees fall into this class: muti 'tree'
- Class III also contains many different types of concepts but most animals fall into this class: embwa 'dog'
- Class IV contains inanimate objects and is the class used for the impersonal 'it': ekitabo 'book'
- Class V contains mainly (but not exclusively) large things and liquids, and can also be used to create augmentatives: ebbeere 'breast', lintu 'giant' (from muntu 'person')
- Class VI contains mainly small things and can be used to create diminutives, adjectival abstract nouns and (in the plural) negative verbal nouns and countries: kabwa 'puppy' (from embwa 'dog' kanafu 'laziness' (from munafu 'lazy'), bukola 'inaction, not to do' (from kukola 'to do, act'), Bungereza 'Britain, England' (from Mungereza 'British, English person')
- Class VII contains many different things including the names of most languages: Oluganda 'Ganda language', Oluzungu 'English language' (from muzungu 'European, white person')
- Class VIII is rarely used but can be used to create pejorative forms: gubwa 'mutt' (from embwa 'dog')
- Class IX is mainly used for infinitives or affirmative verbal nouns: kukola 'action, to do' (from the verb kola 'do, act')
- Class X, which has no singular–plural distinction, is used for mass nouns, usually in the sense of ' drop' or 'precious little': tuzzi 'drop of water' (from mazzi 'water'), tubaka 'sleep'

more complex inflection: Arabic

Form	Past	Meaning	Non-past	Meaning
I	*kataba*	'he wrote'	*yaktubu*	'he writes'
II	*kattaba*	'he made (someone) write'	*yukattibu*	'he makes (someone) write'
III	*kātaba*	'he corresponded with, wrote to (someone)'	*yukātibu*	'"he corresponds with, writes to (someone)'
IV	*'aktaba*	'he dictated'	*yuktibu*	'he dictates'
V	*takattaba*	*nonexistent*	*yatakattabu*	*nonexistent*
VI	*takātaba*	'he corresponded (with someone, esp. mutually)'	*yatakātabu*	'he corresponds (with someone, esp. mutually)'
VII	*inkataba*	'he subscribed'	*yankatibu*	'he subscribes'
VIII	*iktataba*	'he copied'	*yaktatibu*	'he copies'
IX	*iḥmarra*	'he turned red'	*yaḥmarru*	'he turns red'
X	*istaktaba*	'he asked (someone) to write'	*yastaktibu*	'he asks (someone) to write'

forms of *kataba (yaktubu)* 'to write'
Root *ktb*

more complex inflection: Arabic

			Past	Present Indicative	Future Indicative	Subjunctive	Jussive	Long Energetic	Short Energetic	Imperative
Active	Singular									
	1st		katab-t(u) كَتَبْتُ	a-ktub-(u) أَكْتُبُ	sa-'a-ktub-(u) سَأَكْتُبُ	a-ktub-(a) أَكْتُبَ	a-ktub أَكْتُبْ	a-ktub-anna أَكْتُبَنَّ	a-ktub-an أَكْتُبَنْ	– –
	2nd	masculine	katab-t(a) كَتَبْتَ	ta-ktub-(u) تَكْتُبُ	sa-ta-ktub-(u) سَتَكْتُبُ	ta-ktub-(a) تَكْتُبَ	ta-ktub تَكْتُبْ	ta-ktub-anna تَكْتُبَنَّ	ta-ktub-an تَكْتُبَنْ	u-ktub اُكْتُبْ
	2nd	feminine	katab-ti كَتَبْتِ	ta-ktub-īn(a) تَكْتُبِينَ	sa-ta-ktub-īn(a) سَتَكْتُبِينَ	ta-ktub-ī تَكْتُبِي	ta-ktub-ī تَكْتُبِي	ta-ktub-inna تَكْتُبِنَّ	ta-ktub-in تَكْتُبِنْ	u-ktub-ī اُكْتُبِي
	3rd	masculine	katab-(a) كَتَبَ	ya-ktub-(u) يَكْتُبُ	sa-ya-ktub-(u) سَيَكْتُبُ	ya-ktub-(a) يَكْتُبَ	ya-ktub يَكْتُبْ	ya-ktub-anna يَكْتُبَنَّ	ya-ktub-an يَكْتُبَنْ	– –
	3rd	feminine	katab-at كَتَبَتْ	ta-ktub-(u) تَكْتُبُ	sa-ta-ktub-(u) سَتَكْتُبُ	ta-ktub-(a) تَكْتُبَ	ta-ktub تَكْتُبْ	ta-ktub-anna تَكْتُبَنَّ	ta-ktub-an تَكْتُبَنْ	– –
	Dual									
	2nd	masculine & feminine	katab-tumā كَتَبْتُمَا	ta-ktub-ān(i) تَكْتُبَانِ	sa-ta-ktub-ān(i) سَتَكْتُبَانِ	ta-ktub-ā تَكْتُبَا	ta-ktub-ā تَكْتُبَا	ta-ktub-ānni تَكْتُبَانِّ	– –	u-ktub-ā اُكْتُبَا
	3rd	masculine	katab-ā كَتَبَا	ya-ktub-ān(i) يَكْتُبَانِ	sa-ya-ktub-ān(i) سَيَكْتُبَانِ	ya-ktub-ā يَكْتُبَا	ya-ktub-ā يَكْتُبَا	ya-ktub-ānni يَكْتُبَانِّ	– –	– –
	3rd	feminine	katab-atā كَتَبَتَا	ta-ktub-ān(i) تَكْتُبَانِ	sa-ta-ktub-ān(i) سَتَكْتُبَانِ	ta-ktub-ā تَكْتُبَا	ta-ktub-ā تَكْتُبَا	ta-ktub-ānni تَكْتُبَانِّ	– –	– –
	Plural									
	1st		katab-nā كَتَبْنَا	na-ktub-(u) نَكْتُبُ	sa-na-ktub-(u) سَنَكْتُبُ	na-ktub-(a) نَكْتُبَ	na-ktub نَكْتُبْ	na-ktub-anna نَكْتُبَنَّ	na-ktub-an نَكْتُبَنْ	– –
	2nd	masculine	katab-tum كَتَبْتُمْ	ta-ktub-ūn(a) تَكْتُبُونَ	sa-ta-ktub-ūn(a) سَتَكْتُبُونَ	ta-ktub-ū تَكْتُبُوا	ta-ktub-ū تَكْتُبُوا	ta-ktub-unna تَكْتُبُنَّ	ta-ktub-un تَكْتُبُنْ	u-ktub-ū اُكْتُبُوا
	2nd	feminine	katab-tunna كَتَبْتُنَّ	ta-ktub-na تَكْتُبْنَ	sa-ta-ktub-na سَتَكْتُبْنَ	ta-ktub-na تَكْتُبْنَ	ta-ktub-na تَكْتُبْنَ	ta-ktub-nānni تَكْتُبْنَانِّ	– –	u-ktub-na اُكْتُبْنَ
	3rd	masculine	katab-ū كَتَبُوا	ya-ktub-ūn(a) يَكْتُبُونَ	sa-ya-ktub-ūn(a) سَيَكْتُبُونَ	ya-ktub-ū يَكْتُبُوا	ya-ktub-ū يَكْتُبُوا	ya-ktub-unna يَكْتُبُنَّ	ya-ktub-un يَكْتُبُنْ	– –
	3rd	feminine	katab-na كَتَبْنَ	ya-ktub-na يَكْتُبْنَ	sa-ya-ktub-na سَيَكْتُبْنَ	ya-ktub-na يَكْتُبْنَ	ya-ktub-na يَكْتُبْنَ	ya-ktub-nānni يَكْتُبْنَانِّ	– –	– –
Passive	Singular									
	1st		kutib-t(u) كُتِبْتُ	u-ktab-(u) أُكْتَبُ	sa-'u-ktab-(u) سَأُكْتَبُ	u-ktab-(a) أُكْتَبَ	u-ktab أُكْتَبْ	u-ktab-anna أُكْتَبَنَّ	u-ktab-an أُكْتَبَنْ	– –
	2nd	masculine	kutib-t(a) كُتِبْتَ	tu-ktab-(u) تُكْتَبُ	sa-tu-ktab-(u) سَتُكْتَبُ	tu-ktab-(a) تُكْتَبَ	tu-ktab تُكْتَبْ	tu-ktab-anna تُكْتَبَنَّ	tu-ktab-an تُكْتَبَنْ	– –
	2nd	feminine	kutib-ti كُتِبْتِ	tu-ktab-īn(a) تُكْتَبِينَ	sa-tu-ktab-īn(a) سَتُكْتَبِينَ	tu-ktab-ī تُكْتَبِي	tu-ktab-ī تُكْتَبِي	tu-ktab-inna تُكْتَبِنَّ	tu-ktab-in تُكْتَبِنْ	– –
	etc.									
Nominal	Active Participle		kātib كَاتِب			Passive Participle	maktūb مَكْتُوب		Verbal Noun	katb, kitbah, kitābah كَتْب، كِتْبَة، كِتَابَة

forms of *kataba (yaktubu)* 'to write'

Semitic morphology is nonconcatenative

Root	Pattern	Part of Speech	Phonological Form	Orthographic Form	Gloss
ktb	CaCaC	(v)	katav	כתב	‘wrote’
ktb	hiCCiC	(v)	hixtiv	הכתיב	‘dictated’
ktb	miCCaC	(n)	mixtav	מכתב	‘a letter’
ktb	CCaC	(n)	ktav	כתב	‘writing, alphabet’

Inflection by reduplication

anak “child” (Indonesian)
anak-anak “children” (Indonesian)
buah “fruit” (Indonesian)
buah-buahan “various fruits” (Indonesian)

in Indonesian.

Basic Verb	Reduplication	Triplication
kɔul 'to sing'	*kɔukɔul* 'singing'	*kɔukɔukɔul* 'still singing'
mejr 'to sleep'	*mejmejr* 'sleeping'	*mejmejmejr* 'still sleeping'

in Pingelapese.
This process is also productive!

Agreement & inflection in Kayardild

a.	Ngada	kurri-nangku	mala-y.
	1SG.NOM	see-NEG.POTENTIAL	sea-LOC.ACTUAL

'I could not see the sea.' [gyd]

b.	Ngada	kurri-nangku	mala-wu.
	1SG.NOM	see-NEG.POTENTIAL	sea-PROPRIETIVE.FUT

'I won't (be able to) see the sea.' [gyd]

Morphological negation

met numö-ge el-jaqa-te-je
1SG house-LOC NEG-achieve-FUT-INTR.1SG
'I will not reach the house.' [yux]

Kolyma Yukaghir

a. Kim didn't eat the whole pizza.
b. Kim did not eat the whole pizza.

English

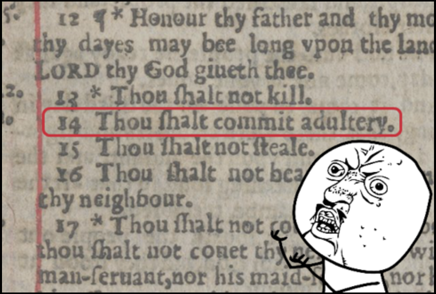
Morphological negation

met numö-ge el-jaqa-te-je
1SG house-LOC NEG-achieve-FUT-INTR.1SG
'I will not reach the house.' [yux]

Kolyma Yukaghir

a. Kim didn't eat the whole pizza.
b. Kim did not eat the whole pizza.

English



Negation matters, as the publishers of the "Wicked Bible" discovered

Derivational morphology

Affix	POS change	Examples
-able	V → A	fixable, doable, understandable
-ive	V → A	assertive, impressive, restrictive
-al	V → N	refusal, disposal, recital
-er	V → N	teacher, worker
-ment	V → N	adjournment, treatment, amazement
-dom	N → N	kingdom, fiefdom
-less	N → A	penniless, brainless
-ic	N → A	cubic, optimistic
-ize	N → V	hospitalize, vaporize
-ize	A → V	modernize, nationalize
-ness	A → N	happiness, sadness
anti-	N → N	antihero, antidepressant
de-	V → V	deactivate, demystify
un-	V → V	untie, unlock, undo
un-	A → A	unhappy, unfair, unintelligent

Compounding and Inflection

ostoskeskuksessa

ostos#keskus+N+Sg+Loc:in

shopping#center+N+Sg+Loc:in

'in the shopping center'

Compounding and Inflection

compounding ostoskeskuksessa

ostos#keskus+N+Sg+Loc:in

shopping#center+N+Sg+Loc:in

'in the shopping center'

Morphological analysis

… is ambiguous

word	прочим
possible analyses	прочий +Adj +Sg +Neut +Instr
	прочий +Adj +Sg +Masc +Instr
	прочий +Adj +Pl +Dat
	прочить +Verb +Pl +1P
	прочее +Pro +Sg +Ins

Generation is (mostly) unambiguous

Writing Systems

morphemes

smallest meaning-bearing unit in a language

- *free morphemes*: can function independently: *dog*, *house*
- *bound morpheme*: appear only as parts of words: *un-* , *-ed*, *-ing*

are morphemes always character sequences?

radicals in Chinese characters can be semantically meaningful

氵	(water)
河	river
湖	lake
海	sea

Consequences for Machine Translation

are open-vocabulary models enough?

- in principle, subword and character-level models can learn morphological generalizations
- in practice, learning morphology from text is hard
  - subword segmentation may not be morphologically sound
  - there may be little surface similarity between related forms

(er) steht	(he) stands
(er) stand	(he) stood

- languages may differ in what information they express morphologically
- there are many good resources for morphological processing:
  - dictionaries with inflection tables
  - lists of stems and affixes
  - rule-based morphological analyzers (finite-state machines)

Morphological Segmentation

- unsupervised segmentation (BPE) crosses morpheme boundaries
- idea: split by morpheme boundary first (prefix, suffixes), then apply BPE [Huck et al., 2017, Pinnis et al., 2017]

BPE	sie alle versch## icken vorsätzlich irreführende Dokumente an Kleinunternehmen in ganz Europa .
suffix + BPE	sie all $$e verschick $$en vorsätz $$lich irreführ $$end $$e Dokument $$e an Kleinunternehm $$en in ganz Europa .
English	they all mail deliberately deceptive documents to small businesses across Europe .

System	test2007		test2008	
	BLEU	TER	BLEU	TER
top 50K voc. (source & target)	25.5	60.9	25.2	60.9
BPE	25.8	60.7	25.6	60.9
compound + BPE	25.9	60.3	25.5	60.6
suffix + BPE	**26.3**	60.0	**26.0**	**60.1**
suffix + compound + BPE	26.2	**59.8**	25.8	60.2
suffix + prefix + compound + BPE	26.1	**59.8**	25.9	60.6
suffix + prefix + compound, 50K	25.9	59.9	25.5	60.3
phrase-based (Huck et al., 2015)	22.6	–	22.1	–

Table 6: English→German experimental results on Europarl (case-sensitive BLEU and TER).

Morphology on Source Side

lemmatized input [Goldwater and McClosky, 2005]

Words:	`Pro někoho by její provedení mělo smysl .`
Lemmas:	`pro někdo být jeho provedení mít smysl .`
Lemmas+Pseudowords:	`pro někdo být PER_3 jeho provedení mít PER_X smysl .`
Modified Lemmas:	`pro někdo být+PER_3 jeho provedení mít+PER_X smysl .`

Figure 2: Various transformations of the Czech sentence from Figure 1. The pseudowords and modified lemmas encode the verb person feature, with the values 3 (third person) and X ("any" person).

Morphology on Source Side

we can easily combine multiple features in NMT [Sennrich and Haddow, 2016]
$\rightarrow$ use word+lemma as input

baseline: only word feature

$$E(stood) = \begin{bmatrix} 0.5 \\ 0.2 \\ 0.3 \\ 0.1 \end{bmatrix}$$

$|F|$ input features

$$E_1(stood) = \begin{bmatrix} 0.4 \\ 0.1 \end{bmatrix} \qquad E_2(stand) = \begin{bmatrix} 0.1 \\ 0.3 \end{bmatrix}$$

$$E_1(stood) \parallel E_2(stand) = \begin{bmatrix} 0.4 \\ 0.1 \\ 0.1 \\ 0.3 \end{bmatrix}$$

Morphology on Target Side

2-step translation [Toutanova et al., 2008]

- predict lemmas in main system
- separate, statistical inflection prediction step

2-step translation in NMT [Tamchyna et al., 2017, García-Martínez et al., 2017]

- predict interleaved lemmas and morph. categories
- inflection generation with finite state transducer

input:	there are a million different kinds of pizza .
baseline:	existují miliony druhů piz@@ zy .
morphgen:	VB-P—3P-AA— existovat NNIP1—–A—- milión NNIP2—–A—- druh NNFS2—–A—- pizza Z:————- .

	baseline	morphgen	Δ
IWSLT	12.89	14.57	1.68
250k	14.87	17.51	2.64
500k	16.96	20.05	**3.09**
1M	18.07	20.95	2.88
2M	20.04	**22.31**	2.27

Inflection Generation with Finite State Transducer

- FSTs can be used to compactly represent morphological grammar
- same transducer can be used for analysis and generation
- cycles allow for elegant modelling of compounding and derivation
- grammars typically hand-designed

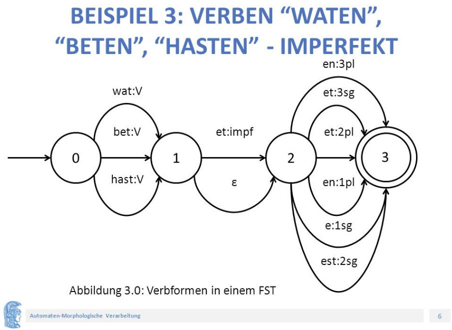


Abbildung 3.0: Verbformen in einem FST

Dr. Christina Alexandris http://slideplayer.gr/slide/10895628/

Neural Inflection Generation

sequence-to-sequence inflection generation [Faruqui et al., 2016]

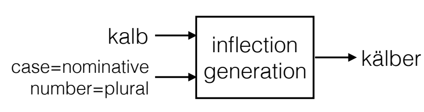


Figure 1: A general inflection generation model.

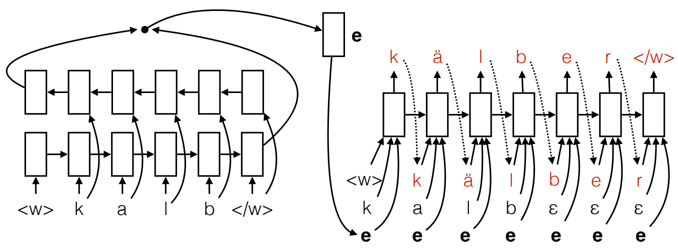


Figure 3: The modified encoder-decoder architecture for inflection generation. Input characters are shown in black and predicted characters are shown in red. • indicates the append operation.

Grammatically Marking Information Missing from Source
Case Study: Controlling Politeness/Formality

T-V distinction

language	informal (T)	formal (V)
Latin	tu	vos
Chinese	你(nǐ)	您 (nín)
French	tu	vous
German	du	Sie
Italian	tu	Lei
Polish	ty	pan
Spanish	tú	usted

Grammatically Marking Information Missing from Source
Case Study: Controlling Politeness/Formality

T-V distinction

language	informal (T)	formal (V)
Latin	tu	vos
Chinese	你(nǐ)	您 (nín)
French	tu	vous
German	du	Sie
Italian	tu	Lei
Polish	ty	pan
Spanish	tú	usted
Early Modern English	thou	ye
Modern English	you	

- inconsistency in T-V choice is a “limitation of MT technology” that is “often frustrat[ing]” to post-editors [Etchegoyhen et al., 2014]

Grammatically Marking Information Missing from Source
Case Study: Controlling Politeness/Formality

T-V distinction

language	informal (T)	formal (V)
Latin	tu	vos
Chinese	你(nǐ)	您 (nín)
French	tu	vous
German	du	Sie
Italian	tu	Lei
Polish	ty	pan
Spanish	tú	usted
Early Modern English	thou	ye
Modern English	you	you

What users want

- inconsistency in T-V choice is a “limitation of MT technology” that is “often frustrat[ing]” to post-editors [Etchegoyhen et al., 2014]

Side constraints [Sennrich et al., 2016]

Core idea

- additional input feature that is based on target-side information
  $\rightarrow$ extra word at end of source sentence
- mark in English text if German translation is polite or not (+noise)

- Are you ok?
- Sind Sie in Ordnung?

- are you ok?
- Bist du in Ordnung?

At test time

- we can control level of politeness by adding side constraints to input

Side constraints [Sennrich et al., 2016]

Core idea

- additional input feature that is based on target-side information
  $\rightarrow$ extra word at end of source sentence
- mark in English text if German translation is polite or not (+noise)

- Are you ok?
- Sind Sie in Ordnung?

- are you ok?
- Bist du in Ordnung?

At test time

- we can control level of politeness by adding side constraints to input

Side constraints [Sennrich et al., 2016]

Core idea

- additional input feature that is based on target-side information
  $\rightarrow$ extra word at end of source sentence
- mark in English text if German translation is polite or not (+noise)

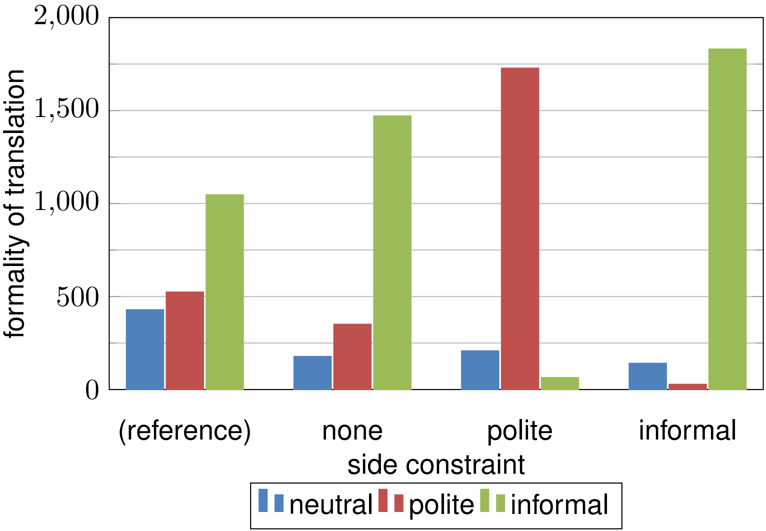
- Are you ok? <polite>
- Sind Sie in Ordnung?

- are you ok? <informal>
- Bist du in Ordnung?

At test time

- we can control level of politeness by adding side constraints to input

Results: politeness as a function of side constraint



Controlling Politeness

- we can effectively control NMT output by providing extra information in input
- here: control politeness
- other applications:
  - control production of other information missing from source text
    - tense
    - evidentiality
    - ...
  - domain adaptation
  - control output language

Evaluating Agreement in Neural MT: Motivation

Evaluating Agreement in Neural MT: Motivation

Kyunghyun Cho
@kchonyc
Following

Fully char-level NMT! It works well on all four language pairs we've considered ({Cs, De, Ru, Fi}->En), and we... fb.me/1oRwyQvZD

RETWEETS 32 LIKES 83

9:12 AM - 11 Oct 2016

2 32 83

Emiel van Miltenburg
@evanmiltenburg
Follow

@kchonyc Are there any benefits to using these models for longer dependencies?

1:16 PM - 11 Oct 2016

1

Evaluating Agreement in Neural MT: Motivation

Kyunghyun Cho @kchonyc — Following

Fully char-level NMT! It works well on all four language pairs we've considered ({Cs, De, Ru, Fi}->En), and we... fb.me/1oRwyQvZD

RETWEETS 32 LIKES 83

9:12 AM - 11 Oct 2016

2 32 83

Emiel van Miltenburg @evanmiltenburg — Follow

@kchonyc Are there any benefits to using these models for longer dependencies?

1:16 PM - 11 Oct 2016

1

Kyunghyun Cho @kchonyc — Following

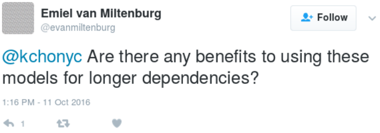
@evanmiltenburg ah well that's a difficult question!

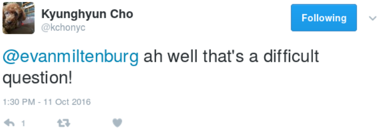
1:30 PM - 11 Oct 2016

1

Evaluating Agreement in Neural MT: Motivation


Kyunghyun Cho
@kchonyc
Following
Fully char-level NMT! It works well on all four language pairs we've considered ({Cs, De, Ru, Fi}->En), and we... fb.me/1oRwyQvZD
RETWEETS 32 LIKES 83
9:12 AM - 11 Oct 2016
2 32 83


Emiel van Miltenburg
@evanmiltenburg
Follow
@kchonyc Are there any benefits to using these models for longer dependencies?
1:16 PM - 11 Oct 2016
1


Kyunghyun Cho
@kchonyc
Following
@evanmiltenburg ah well that's a difficult question!
1:30 PM - 11 Oct 2016
1

text representation

word-level	but as the **example** of Mobilking in Poland **shows** |—— 5 steps ——|
subword-level (byte-pair encoding)	but as the **example** of Mobil+ king in Poland **shows** |—— 6 steps ——|
character-level	b u t _ a s _ t h e _ **e x a m p l e** _ o f _ M o b i l k i n g _ i n _ P o l a n d _ **s h o w s** |—— 29 steps ——|

Evaluating Agreement in Neural MT: Motivation

Kyunghyun Cho @kchonyc — Following

Fully char-level NMT! It works well on all four language pairs we've considered ({Cs, De, Ru, Fi}->En), and we... fb.me/1oRwyQvZD

RETWEETS 32 LIKES 83

9:12 AM - 11 Oct 2016

2 32 83

Emiel van Miltenburg @evanmiltenburg — Follow

@kchonyc Are there any benefits to using these models for longer dependencies?

1:16 PM - 11 Oct 2016

1

Kyunghyun Cho @kchonyc — Following

@evanmiltenburg ah well that's a difficult question!

1:30 PM - 11 Oct 2016

1

text representation

word-level	but as the **example** of UNK in Poland **shows** |—— 5 steps ——|
subword-level (byte-pair encoding)	but as the **example** of Mobil+ king in Poland **shows** |—— 6 steps ——|
character-level	b u t _ a s _ t h e _ **e x a m p l e** _ o f _ M o b i l k i n g _ i n _ P o l a n d _ **s h o w s** |—— 29 steps ——|

How to Assess Specific Aspects in MT?

- human evaluation
  - × costly; hard to compare to previous work
- automatic metrics (BLEU)
  - × too coarse; blind towards specific aspects

How to Assess Specific Aspects in MT?

- human evaluation
  $\times$ costly; hard to compare to previous work
- automatic metrics (BLEU)
  $\times$ too coarse; blind towards specific aspects

contrastive translation pairs

- NMT models assign probability to any translation
- binary classification task: which translation is better?
- choice between reference translation and contrastive variant
  $\rightarrow$ corrupted with single error of specific type
- $\approx$ minimal pairs in linguistics

Assessment with Contrastive Translation Pairs

workflow

- researcher wants to analyse difficult translation problem
- researcher predicts what errors NMT system might make
- researcher creates test set with correct translations and corrupted variants
- test set allows automatic, quantitative, and reproducible analysis of NMT model

example

Assessment with Contrastive Translation Pairs

workflow

- researcher wants to analyse difficult translation problem
- researcher predicts what errors NMT system might make
- researcher creates test set with correct translations and corrupted variants
- test set allows automatic, quantitative, and reproducible analysis of NMT model

example

- subject–verb agreement

Assessment with Contrastive Translation Pairs

workflow

- researcher wants to analyse difficult translation problem
- researcher predicts what errors NMT system might make
- researcher creates test set with correct translations and corrupted variants
- test set allows automatic, quantitative, and reproducible analysis of NMT model

example

- subject–verb agreement
- change grammatical number of verb to introduce agreement error

Assessment with Contrastive Translation Pairs

workflow

- researcher wants to analyse difficult translation problem
- researcher predicts what errors NMT system might make
- researcher creates test set with correct translations and corrupted variants
- test set allows automatic, quantitative, and reproducible analysis of NMT model

example

- subject–verb agreement
- change grammatical number of verb to introduce agreement error
- 35000 contrastive pairs created with simple linguistic rules

Contrastive Translation Pairs

	sentence	prob.
English	[...] that the **plan will** be approved	
German (correct)	[...], dass der **Plan** verabschiedet **wird**	0.1 ✓
German (contrastive)	* [...], dass der **Plan** verabschiedet **werden**	0.01

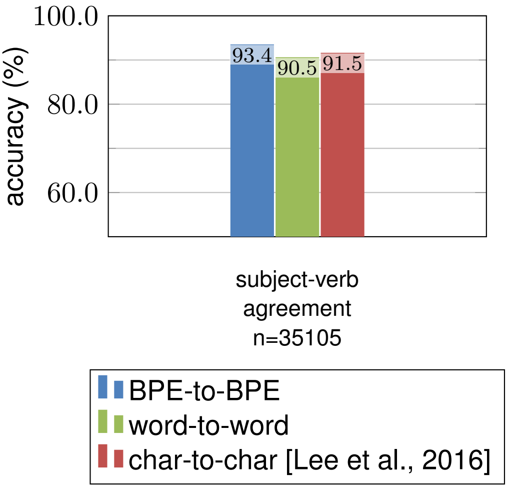
subject-verb agreement

LingEval97: A Test Set of Contrastive Translation Pairs

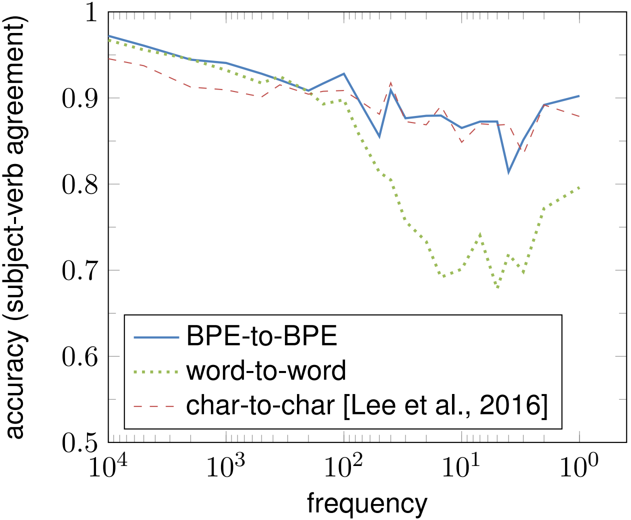
LingEval97

- 97 000 contrastive translation pairs
- based on English→German WMT test sets
- rule-based, automatic creation of errors
- 7 error types
- metadata for in-depth analysis:
  - error type
  - distance between words
  - word frequency in WMT15 training set

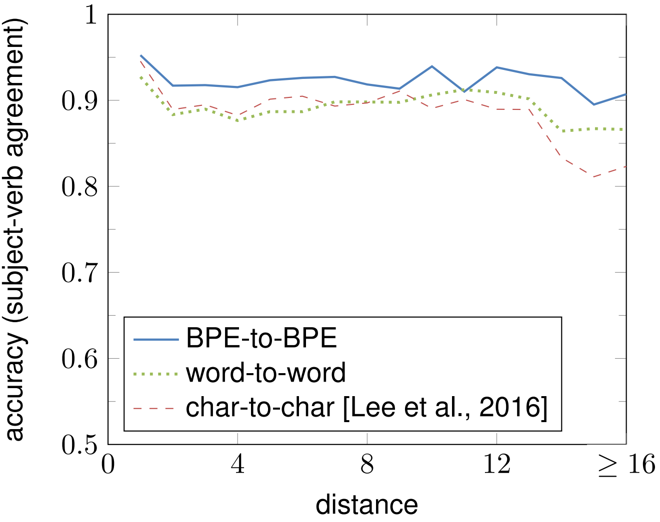
Results: Text Representation



Results: Text Representation



Results: Text Representation



What Did We Learn?

- word-level model is poor for rare words
- character-level model is poor for long distances
- BPE subword segmentation is good compromise

Conclusions

- morphology in NMT (and NLP) still very much open research area
- be aware of morphological properties of languages that you work with
  - best word segmentation strategy may be language-dependent (word segmentation for non-concatenative morphology?)
  - morphological simplification may help translation
  - languages differ in what information is expressed grammatically
- agreement, a traditionally hard problem for MT, is solved relatively well by NMT

Further Reading

- Bender, chapters 2–4

Bibliography I

Etchegoyhen, T., Bywood, L., Fishel, M., Georgakopoulou, P., Jiang, J., Loenhout, G. V., Pozo, A. D., Maucec, M. S., Turner, A., and Volk, M. (2014).
Machine Translation for Subtitling: A Large-Scale Evaluation.
In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik, Iceland. European Language Resources Association (ELRA).

Faruqui, M., Tsvetkov, Y., Neubig, G., and Dyer, C. (2016).
Morphological Inflection Generation Using Character Sequence to Sequence Learning.
In
Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Langua
pages 634–643, San Diego, California. Association for Computational Linguistics.

García-Martínez, M., Barrault, L., and Bougares, F. (2017).
Neural Machine Translation by Generating Multiple Linguistic Factors.
In
Statistical Language and Speech Processing - 5th International Conference, SLSP 2017, Le Mans, France, October 23-25, 2017, Pro
pages 21–31.

Goldwater, S. and McClosky, D. (2005).
Improving Statistical MT through Morphological Analysis.
In
Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing,
pages 676–683, Vancouver, British Columbia, Canada. Association for Computational Linguistics.

Huck, M., Riess, S., and Fraser, A. (2017).
Target-side Word Segmentation Strategies for Neural Machine Translation.
In Proceedings of the Second Conference on Machine Translation, Volume 1: Research Papers, Copenhagen, Denmark. Association for Computational Linguistics.

Bibliography II

Lee, J., Cho, K., and Hofmann, T. (2016).
Fully Character-Level Neural Machine Translation without Explicit Segmentation.
ArXiv e-prints.

Pinnis, M., Krislauks, R., Deksne, D., and Miks, T. (2017).
Neural Machine Translation for Morphologically Rich Languages with Improved Sub-word Units and Synthetic Data.
In Text, Speech, and Dialogue - 20th International Conference, TSD 2017, pages 237–245, Prague, Czech Republic.

Sennrich, R. and Haddow, B. (2016).
Linguistic Input Features Improve Neural Machine Translation.
In Proceedings of the First Conference on Machine Translation, Volume 1: Research Papers, pages 83–91, Berlin, Germany.

Sennrich, R., Haddow, B., and Birch, A. (2016).
Controlling Politeness in Neural Machine Translation via Side Constraints.
In
Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Langua
pages 35–40, San Diego, California.

Tamchyna, A., Weller-Di Marco, M., and Fraser, A. (2017).
Modeling Target-Side Inflection in Neural Machine Translation.
In Second Conference on Machine Translation (WMT17).

Toutanova, K., Suzuki, H., and Ruopp, A. (2008).
Applying Morphology Generation Models to Machine Translation.
In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics.