

The secret life of words

Example

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

Pair	Probability
P(<i>Häuser</i> <i>houses</i>)	0.22393465
P(<i>nach</i> <i>houses</i>)	0.02897585
P(<i>Daches</i> <i>houses</i>)	0.02893843
P(<i>halten</i> <i>houses</i>)	0.02756981
P(<i>beihilfe</i> <i>fähig</i> <i>houses</i>)	0.02536119
P(<i>Wohnungen</i> <i>houses</i>)	0.02536087
P(<i>Privathäuser</i> <i>houses</i>)	0.02513098
P(<i>EFRE – Verordnung</i> <i>houses</i>)	0.02480666
P(<i>immer</i> <i>houses</i>)	0.02395412
P(<i>Palästinenser</i> <i>fahrer</i> <i>houses</i>)	0.02318363
P(<i>Parlamentsreden</i> <i>houses</i>)	0.02318363
P(<i>abzureißen</i> <i>houses</i>)	0.02318363
P(<i>gleichmachen</i> <i>houses</i>)	0.02318363
P(<i>Erdboden</i> <i>houses</i>)	0.02318363
P(<i>bombardieren</i> <i>houses</i>)	0.02318363
P(<i>genauen</i> <i>houses</i>)	0.0231831
P(<i>schießen</i> <i>houses</i>)	0.0231829
P(<i>anderem</i> <i>houses</i>)	0.02315655
P(<i>fest</i> <i>houses</i>)	0.0222202
P(<i>gezielt</i> <i>housesng</i>)	0.02200406

Example

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

Pair	Probability
$P(Häuser houses)$	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
$P(Haus house)$	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

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

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
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
$P(Haus house)$	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
number

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 ...

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

Quick note: gender/ class

- German: masculine, feminine, neuter

	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

Quick note: gender/ class

- German: masculine, feminine, neuter
- Wayamba: masculine, feminine, neuter, vegetable

	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	<i>kataba</i>	'he wrote'	<i>yaktubu</i>	'he writes'
II	<i>kattaba</i>	'he made (someone) write'	<i>yukattibu</i>	'he makes (someone) write'
III	<i>kātaba</i>	'he corresponded with, wrote to (someone)'	<i>yukātibu</i>	'he corresponds with, writes to (someone)'
IV	<i>ʾaktaba</i>	'he dictated'	<i>yuktibu</i>	'he dictates'
V	<i>takattaba</i>	<i>nonexistent</i>	<i>yatakattabu</i>	<i>nonexistent</i>
VI	<i>takātaba</i>	'he corresponded (with someone, esp. mutually)'	<i>yatakātabu</i>	'he corresponds (with someone, esp. mutually)'
VII	<i>inkataba</i>	'he subscribed'	<i>yankatibu</i>	'he subscribes'
VIII	<i>iktataba</i>	'he copied'	<i>yaktatibu</i>	'he copies'
IX	<i>iḥmarra</i>	'he turned red'	<i>yaḥmarru</i>	'he turns red'
X	<i>istaktaba</i>	'he asked (someone) to write'	<i>yastaktibu</i>	'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) كَاتَبْتُ	a-ktub-(u) اَكْتُبْ	sa-'a-ktub-(u) سَأَكْتُبُ	a-ktub-(a) اَكْتُبْ	a-ktub اَكْتُبْ	a-ktub-anna اَكْتُبْ اِنَّا	a-ktub-an اَكْتُبْ اَنْ	—
	2nd	masculine	katab-t(a) كَاتَبْتَ	ta-ktub-(u) تَكْتُبْ	sa-ta-ktub-(u) سَأَتَكْتُبُ	ta-ktub-(a) تَكْتُبْ	ta-ktub تَكْتُبْ	ta-ktub-anna تَكْتُبْ اِنَّا	ta-ktub-an تَكْتُبْ اَنْ	u-ktub اَكْتُبْ
		feminine	katab-ti كَاتَبْتِ	ta-ktub-īn(a) تَكْتُبِينَ	sa-ta-ktub-īn(a) سَأَتَكْتُبِينَ	ta-ktub-ī تَكْتُبِي	ta-ktub-ī تَكْتُبِي	ta-ktub-inna تَكْتُبِي اِنَّا	ta-ktub-in تَكْتُبِي اَنْ	u-ktub-ī اَكْتُبِي
	3rd	masculine	katab-(a) كَاتَبَ	ya-ktub-(u) يَكْتُبْ	sa-ya-ktub-(u) سَيَكْتُبُ	ya-ktub-(a) يَكْتُبْ	ya-ktub يَكْتُبْ	ya-ktub-anna يَكْتُبْ اِنَّا	ya-ktub-an يَكْتُبْ اَنْ	—
		feminine	katab-at كَاتَبَتْ	ta-ktub-(u) تَكْتُبْ	sa-ta-ktub-(u) سَأَتَكْتُبْ	ta-ktub-(a) تَكْتُبْ	ta-ktub تَكْتُبْ	ta-ktub-anna تَكْتُبْ اِنَّا	ta-ktub-an تَكْتُبْ اَنْ	—
	Dual									
	2nd	masculine & feminine	katab-tumā كَاتَبْتُمَا	ta-ktub-ān(i) تَكْتُبَانِ	sa-ta-ktub-ān(i) سَأَتَكْتُبَانِ	ta-ktub-ā تَكْتُبَا	ta-ktub-ā تَكْتُبَا	ta-ktub-ānni تَكْتُبَا اِنَّا	—	u-ktub-ā اَكْتُبَا
		masculine	katab-ā كَاتَبَا	ya-ktub-ān(i) يَكْتُبَانِ	sa-ya-ktub-ān(i) سَيَكْتُبَانِ	ya-ktub-ā يَكْتُبَا	ya-ktub-ā يَكْتُبَا	ya-ktub-ānni يَكْتُبَا اِنَّا	—	—
	3rd	feminine	katab-atā كَاتَبَتَا	ta-ktub-ān(i) تَكْتُبَانِ	sa-ta-ktub-ān(i) سَأَتَكْتُبَانِ	ta-ktub-ā تَكْتُبَا	ta-ktub-ā تَكْتُبَا	ta-ktub-ānni تَكْتُبَا اِنَّا	—	—
	Plural									
	1st		katab-nā كَاتَبْنَا	na-ktub-(u) نَكْتُبْ	sa-na-ktub-(u) سَنَكْتُبُ	na-ktub-(a) نَكْتُبْ	na-ktub نَكْتُبْ	na-ktub-anna نَكْتُبْ اِنَّا	na-ktub-an نَكْتُبْ اَنْ	—
	2nd	masculine	katab-tum كَاتَبْتُمْ	ta-ktub-ūn(a) تَكْتُبُونَ	sa-ta-ktub-ūn(a) سَأَتَكْتُبُونَ	ta-ktub-ū تَكْتُبُوا	ta-ktub-ū تَكْتُبُوا	ta-ktub-unna تَكْتُبُوا اِنَّا	ta-ktub-un تَكْتُبُوا اَنْ	u-ktub-ū اَكْتُبُوا
		feminine	katab-tunna كَاتَبْتُنَّ	ta-ktub-na تَكْتُبْنَ	sa-ta-ktub-na سَأَتَكْتُبْنَ	ta-ktub-na تَكْتُبْنَ	ta-ktub-na تَكْتُبْنَ	ta-ktub-nānni تَكْتُبْنَ اِنَّا	—	u-ktub-na اَكْتُبْنَ
	3rd	masculine	katab-ū كَاتَبُوا	ya-ktub-ūn(a) يَكْتُبُونَ	sa-ya-ktub-ūn(a) سَيَكْتُبُونَ	ya-ktub-ū يَكْتُبُوا	ya-ktub-ū يَكْتُبُوا	ya-ktub-unna يَكْتُبُوا اِنَّا	ya-ktub-un يَكْتُبُوا اَنْ	—
		feminine	katab-na كَاتَبْنَ	ya-ktub-na يَكْتُبْنَ	sa-ya-ktub-na سَيَكْتُبْنَ	ya-ktub-na يَكْتُبْنَ	ya-ktub-na يَكْتُبْنَ	ya-ktub-nānni يَكْتُبْنَ اِنَّا	—	—
Passive	Singular									
	1st		kutib-t(u) كُتِبْتُ	u-ktab-(u) اُكْتُبْ	sa-'u-ktab-(u) سَأُكْتُبُ	u-ktab-(a) اُكْتُبْ	u-ktab اُكْتُبْ	u-ktab-anna اُكْتُبْ اِنَّا	u-ktab-an اُكْتُبْ اَنْ	—
	2nd	masculine	kutib-t(a) كُتِبْتَ	tu-ktab-(u) تُكْتُبْ	sa-tu-ktab-(u) سَأَتُكْتُبُ	tu-ktab-(a) تُكْتُبْ	tu-ktab تُكْتُبْ	tu-ktab-anna تُكْتُبْ اِنَّا	tu-ktab-an تُكْتُبْ اَنْ	—
		feminine	kutib-ti كُتِبْتِ	tu-ktab-īn(a) تُكْتُبِينَ	sa-tu-ktab-īn(a) سَأَتُكْتُبِينَ	tu-ktab-ī تُكْتُبِي	tu-ktab-ī تُكْتُبِي	tu-ktab-inna تُكْتُبِي اِنَّا	tu-ktab-in تُكْتُبِي اَنْ	—
	etc.									
Nominal	Active Participle				Passive Participle				Verbal Noun	
	kātib كَاتِبٌ				maktūb مَكْتُوبٌ				katb, kitbah, kitābah كَتَبَ، كَتَبَتْ، كَتَبُوا	

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
<i>kɔul</i> 'to sing'	<i>kɔukɔul</i> 'singing'	<i>kɔukɔukɔul</i> 'still singing'
<i>mejɾ</i> 'to sleep'	<i>mejmejɾ</i> 'sleeping'	<i>mejmejmejɾ</i> '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

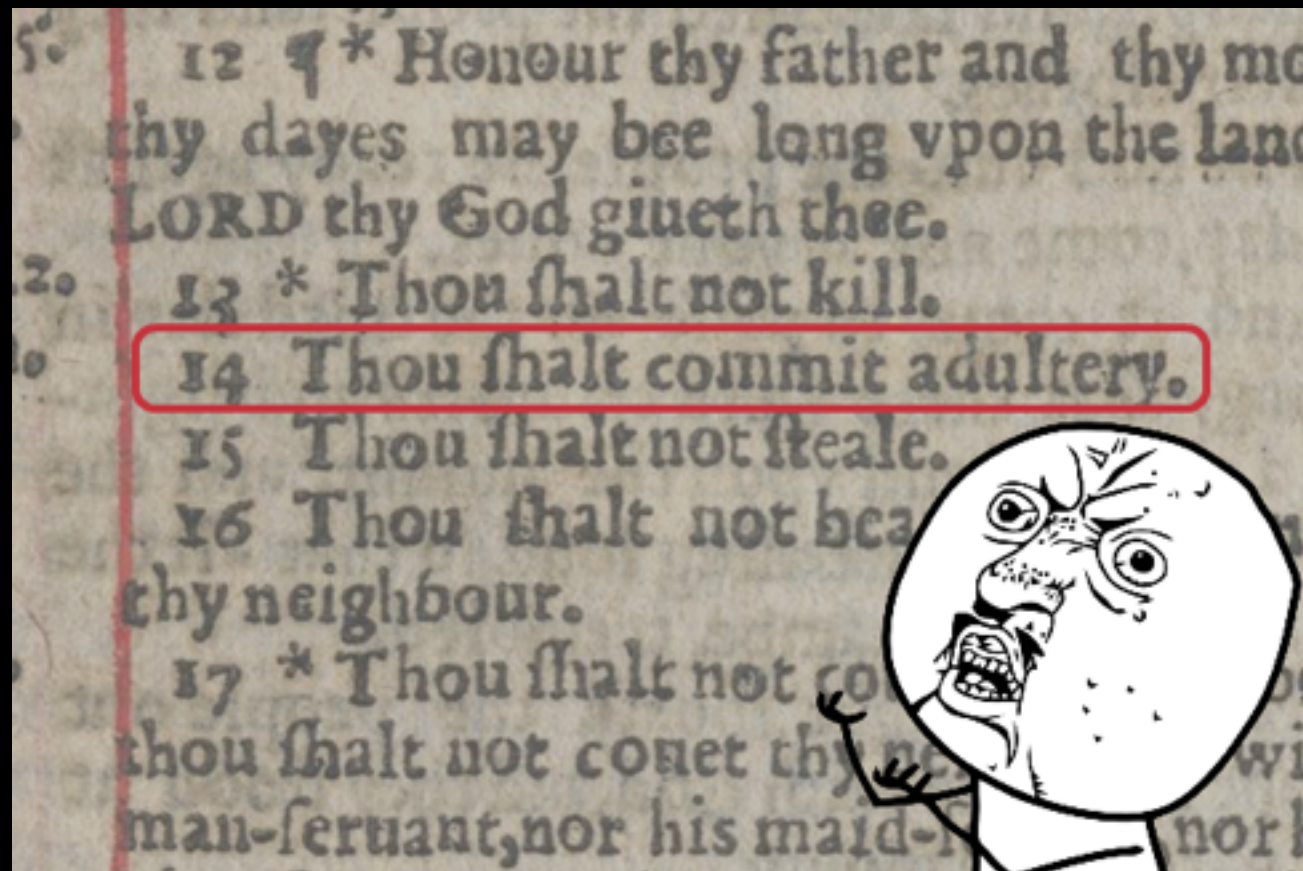
Morphological negation

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

Translating *from* rich morphology

Words:

Pro někoho by její provedení mělo smysl .

Translating *from* rich morphology

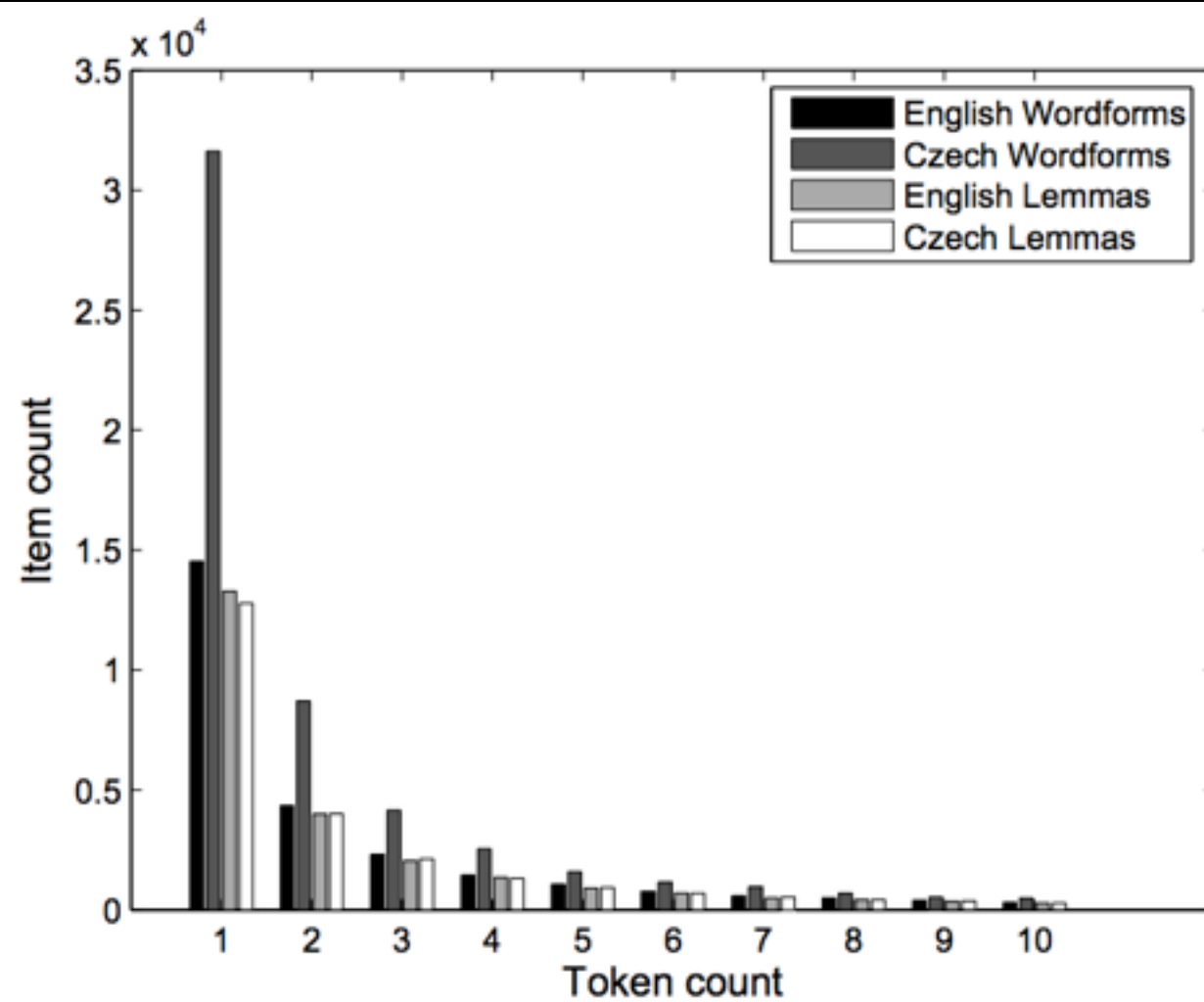
Words:	Pro někoho by její provedení mělo smysl .
Lemmas:	pro někdo být jeho provedení mít smysl .

Translating *from* rich morphology

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 .

Translating *from* rich morphology

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 .



Translating *from* rich morphology

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 .

	Dev	Test
word-to-word	.311	.270
lemmatize all	.355	.299
except Pro	.350	
except Pro, V, N	.346	
lemmatize $n < 50$.370	.306
truncate all	.353	.283

Translating *from* rich morphology

Rather than decide on a particular representation of words,
allow translation system to choose.

1	2	3	4	5	6	7	8	9	10	11	12
z	amerického americký	břehu břeh	atlantiku atlantik	se s	veskerá	taková takový	odůvodnění	jeví jevit	jako	naprosto	bizarní

Translating *from* rich morphology

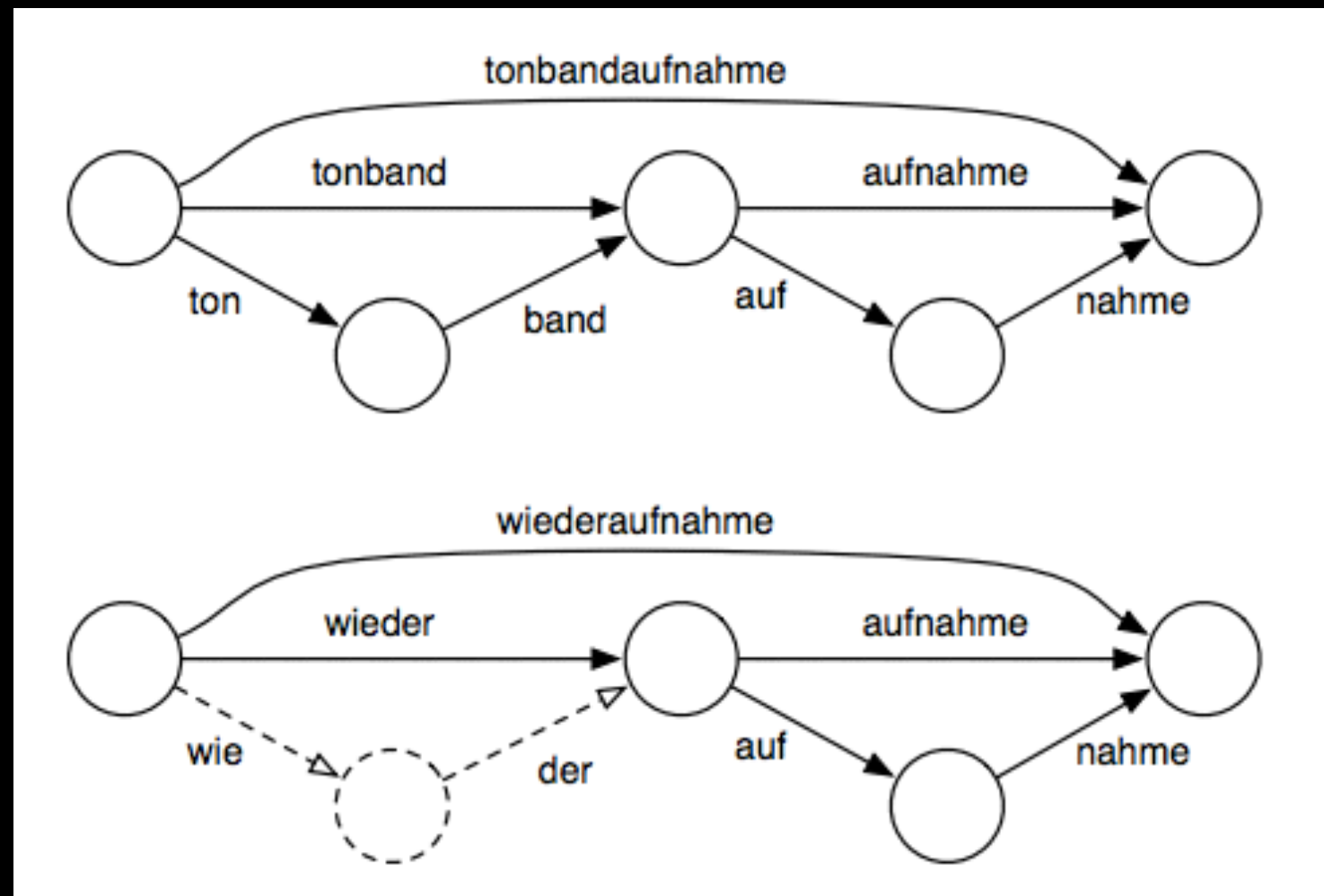
Rather than decide on a particular representation of words,
allow translation system to choose.

1	2	3	4	5	6	7	8	9	10	11	12
z	amerického americký	břehu břeh	atlantiku atlantik	se s	veskerá	taková takový	odůvodnění	jeví jevit	jako	naprosto	bizarní

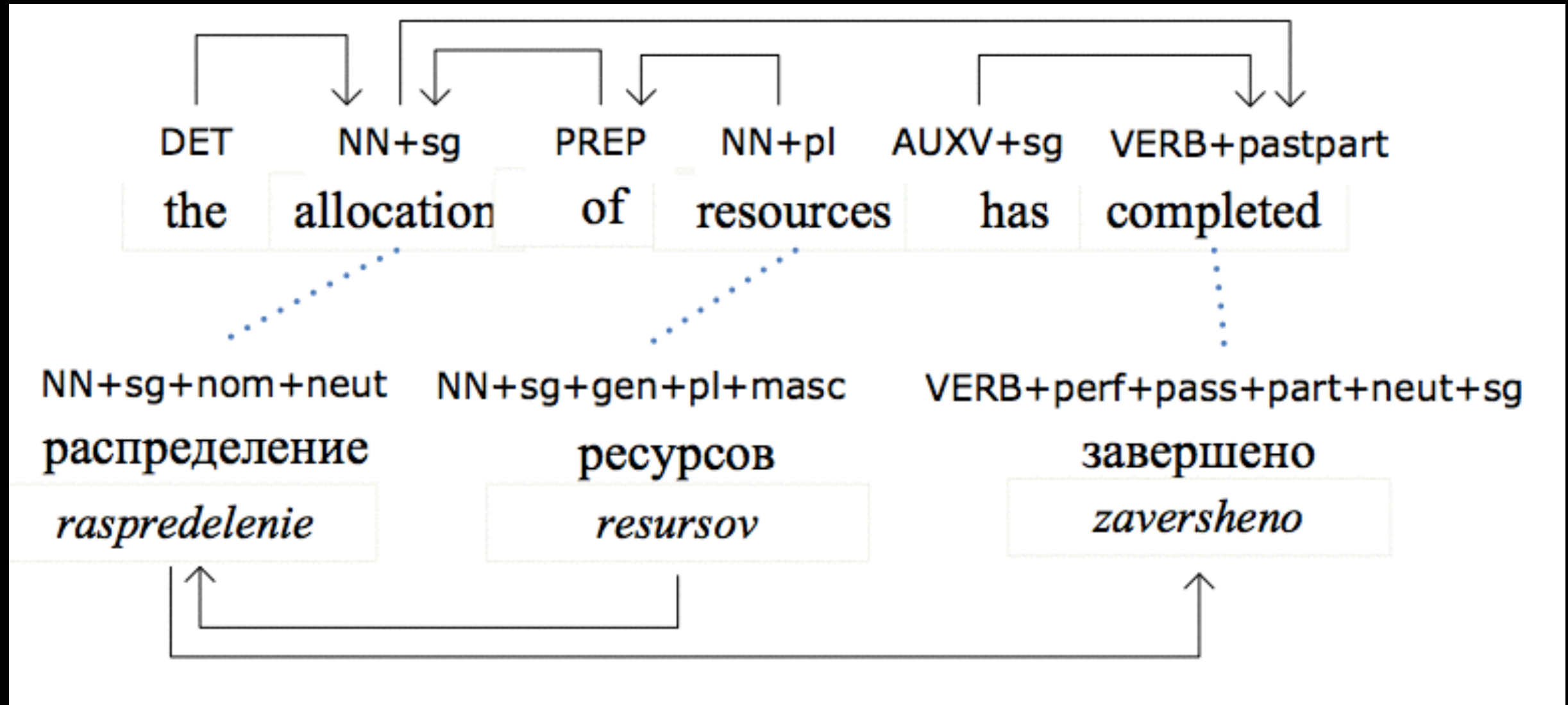
Input	BLEU	Sample translation
SURFACE	22.74	From the US side of the Atlantic all such odůvodnění appears to be a totally bizarre.
LEMMA	22.50	From the side of the Atlantic with any such justification seem completely bizarre.
TRUNC ($l=6$)	22.07	From the bank of the Atlantic, all such justification appears to be totally bizarre.
backoff (SURFACE+LEMMA)	23.94	From the US bank of the Atlantic, all such justification appears to be totally bizarre.
CN (SURFACE+LEMMA)	25.01	From the US side of the Atlantic all such justification appears to be a totally bizarre.
CN (SURFACE+TRUNC)	23.57	From the US Atlantic any such justification appears to be a totally bizarre.

Translating *from* rich morphology

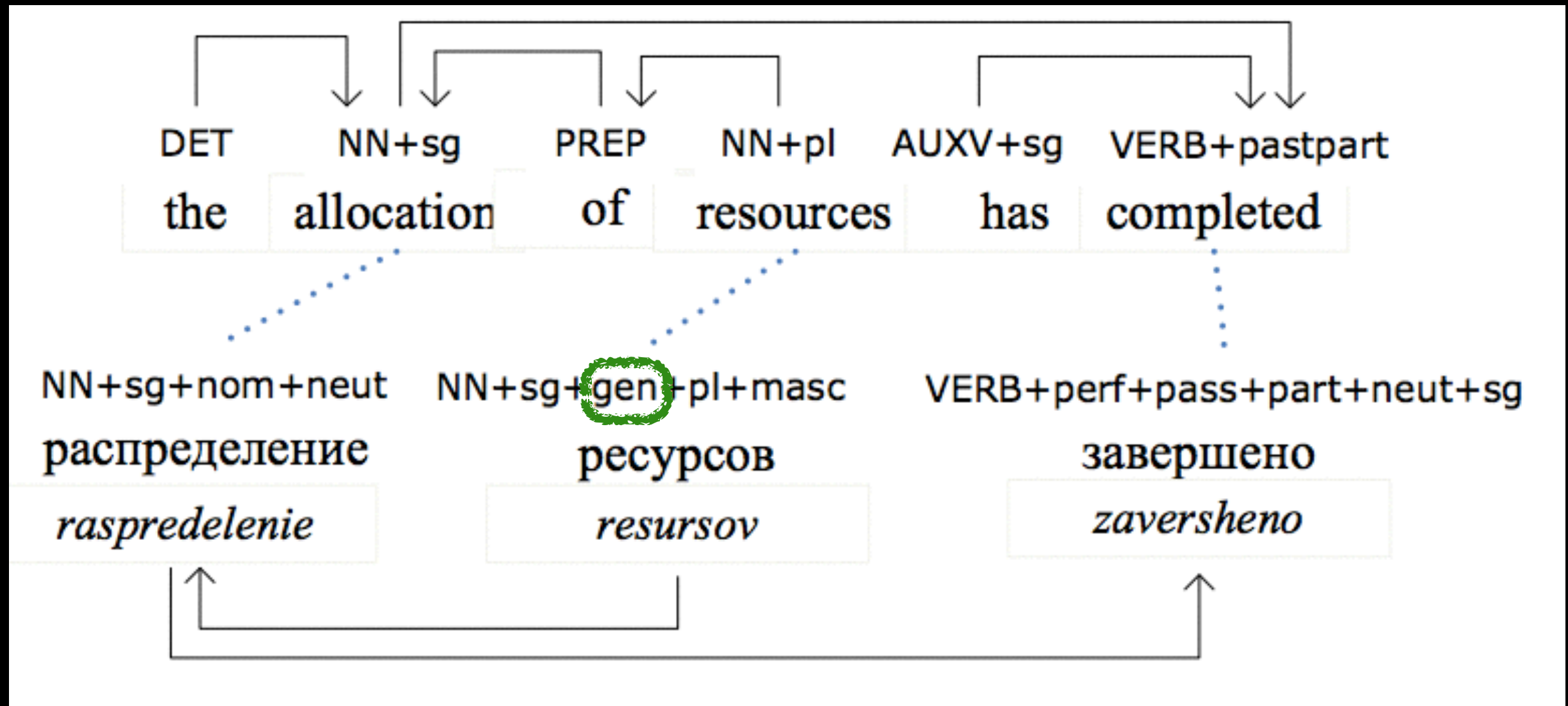
Related: uncompounding



Translating *to* rich morphology

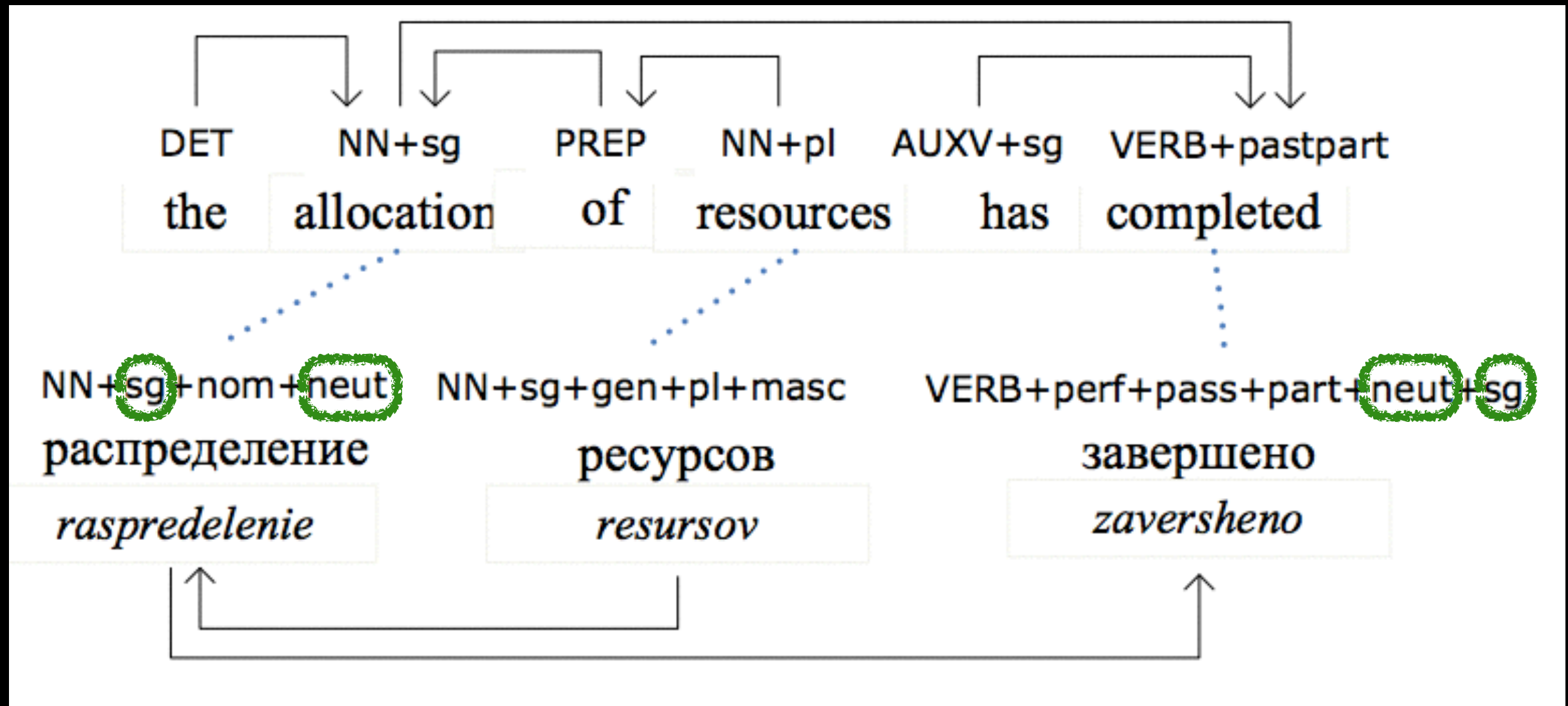


Translating *to* rich morphology



genitive case: marks noun modifying another noun

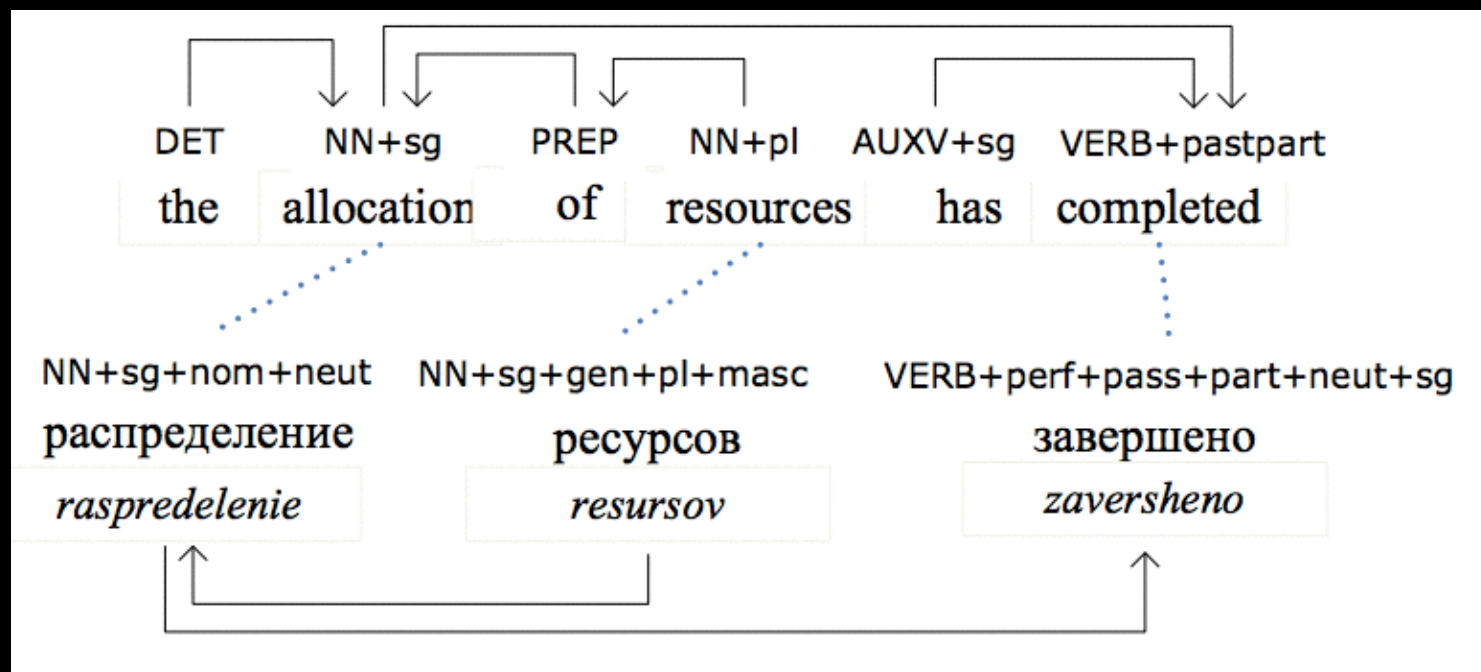
Translating *to* rich morphology



verb and its argument must agree on these features

Translating *to* rich morphology

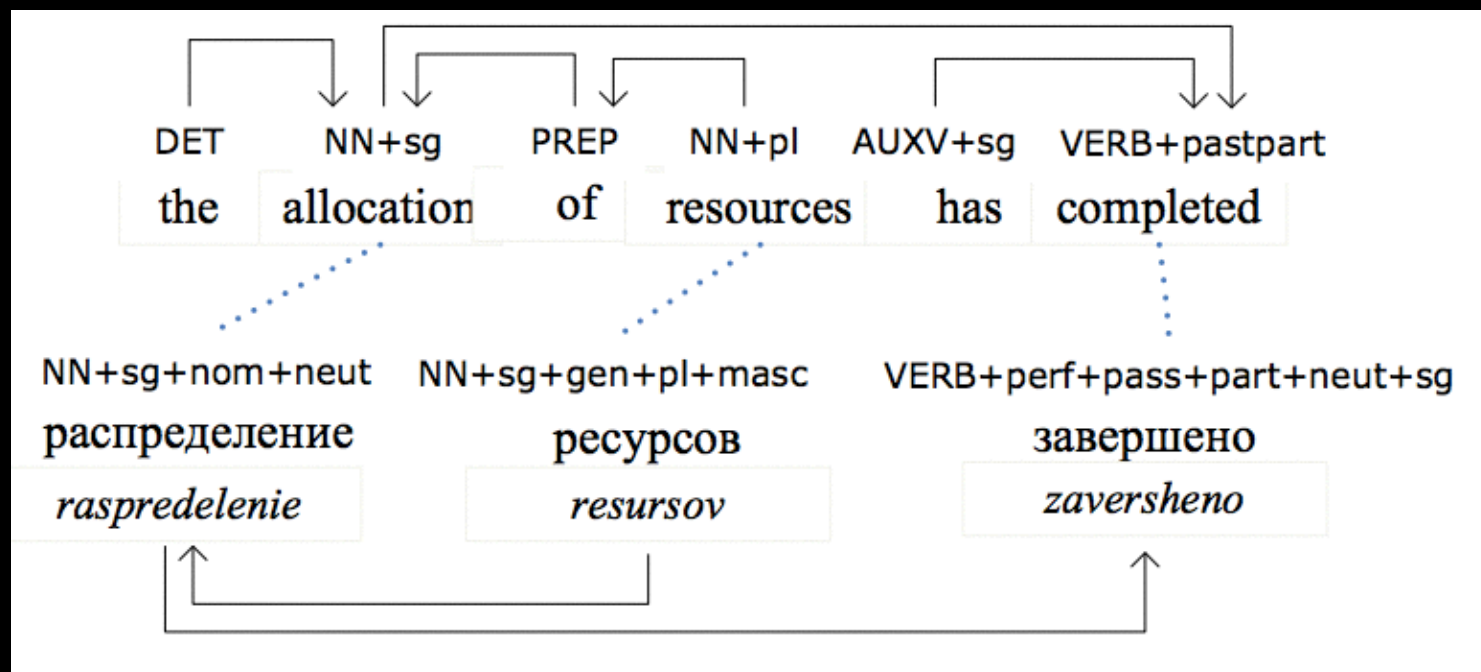
Feature categories	Instantiations
Monolingual lexical	
Word stem	$s_{t-1}, s_{t-2}, s_t, s_{t+1}$
Predicted word	y_t, y_{t-1}, y_{t-2}
Monolingual morphological	
f : POS, Person, Number, Gender, Tense Neg, Det, Prep, Conj, ObjPron, PossPron	$f(y_{t-2}), f(y_{t-1}), f(y_t)$
Monolingual syntactic	
Parent stem	$s_{HEAD(t)}$
Bilingual lexical	
Aligned word set Al	Al_t, Al_{t-1}, Al_{t+1}
Bilingual morph & syntactic	
f : POS, Person, Number, Gender, Tense Neg, Det, Prep, Conj, ObjPron, PossPron, Comp	$f(Al_t), f(Al_{t-1}),$ $f(Al_{t+1}), f(Al_{HEAD(t)})$



$$p(y|x) = \prod_{t=1}^n p(y_t | y_{t-1} \dots y_{t-k}, x_t)$$

Translating *to* rich morphology

Feature categories	Instantiations
Monolingual lexical	
Word stem	$s_{t-1}, s_{t-2}, s_t, s_{t+1}$
Predicted word	y_t, y_{t-1}, y_{t-2}
Monolingual morphological	
f : POS, Person, Number, Gender, Tense Neg, Det, Prep, Conj, ObjPron, PossPron	$f(y_{t-2}), f(y_{t-1}), f(y_t)$
Monolingual syntactic	
Parent stem	$s_{HEAD(t)}$
Bilingual lexical	
Aligned word set Al	Al_t, Al_{t-1}, Al_{t+1}
Bilingual morph & syntactic	
f : POS, Person, Number, Gender, Tense Neg, Det, Prep, Conj, ObjPron, PossPron, Comp	$f(Al_t), f(Al_{t-1}),$ $f(Al_{t+1}), f(Al_{HEAD(t)})$

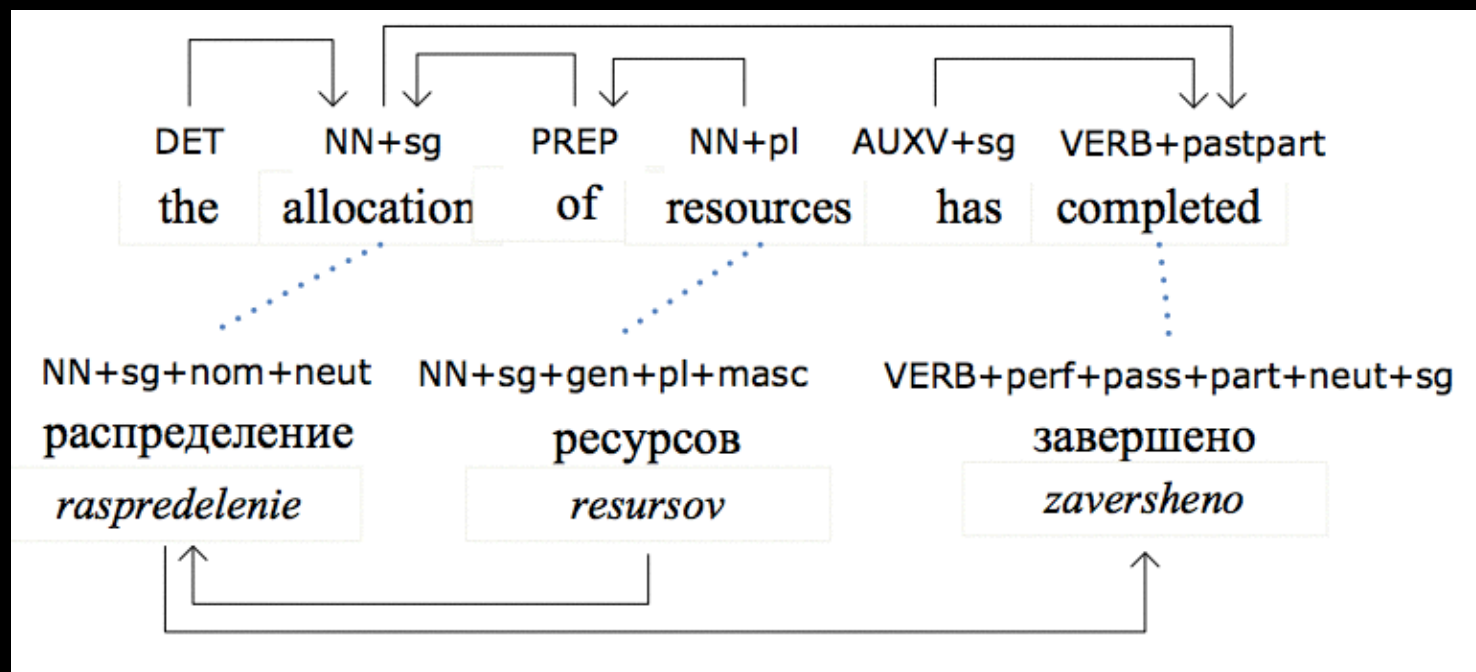


$$p(y|x) = \prod_{t=1}^n p(y_t | y_{t-1} \dots y_{t-k}, x_t)$$

Translating *to* rich morphology

Model	BLEU
Base MT ($n=1$)	29.24
Method 1 ($n=1$)	30.44
Method 1 ($n=16$)	30.61
Method 2 ($n=1$)	30.79
Method 2 ($n=16$)	31.24
Method 3 ($n=1$)	31.42
Method 3 ($n=32$)	31.80

Feature categories	Instantiations
Monolingual lexical	
Word stem	$s_{t-1}, s_{t-2}, s_t, s_{t+1}$
Predicted word	y_t, y_{t-1}, y_{t-2}
Monolingual morphological	
f : POS, Person, Number, Gender, Tense	$f(y_{t-2}), f(y_{t-1}), f(y_t)$
Neg, Det, Prep, Conj, ObjPron, PossPron	
Monolingual syntactic	
Parent stem	$s_{HEAD(t)}$
Bilingual lexical	
Aligned word set Al	Al_t, Al_{t-1}, Al_{t+1}
Bilingual morph & syntactic	
f : POS, Person, Number, Gender, Tense	$f(Al_t), f(Al_{t-1}),$
Neg, Det, Prep, Conj, ObjPron, PossPron,	$f(Al_{t+1}), f(Al_{HEAD(t)})$
Comp	



$$p(y|x) = \prod_{t=1}^n p(y_t | y_{t-1} \dots y_{t-k}, x_t)$$

Translating *to* rich morphology

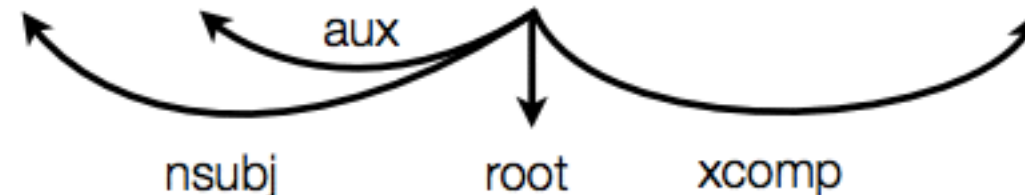
σ:пытаться_V + μ:mis-sfm-e

она **пыталась** пересечь пути на ее велосипеде

she had attempted to cross the road on her bike

C50 C473 C28 C8 C275 C37 C43 C82 C94 C331

PRP VBD VBN TO VB DT NN IN PRP\$ NN



Translating *to* rich morphology

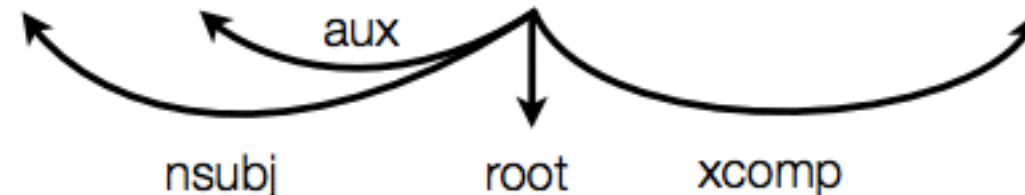
σ:пытаться_V + μ:mis-sfm-e

она **пыталась** пересечь пути на ее велосипеде

she had attempted to cross the road on her bike

C50 C473 C28 C8 C275 C37 C43 C82 C94 C331

PRP VBD VBN TO VB DT NN IN PRP\$ NN



	EN→RU	EN→HE	EN→SW
Baseline	14.7±0.1	15.8±0.3	18.3±0.1
+Class LM	15.7±0.1	16.8±0.4	18.7±0.2
+Synthetic			
unsupervised	16.2±0.1	17.6±0.1	19.0±0.1
supervised	16.7±0.1	—	—