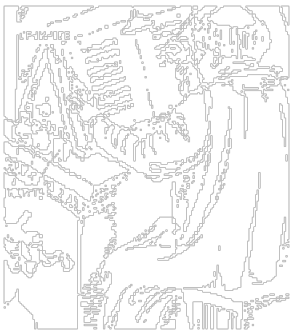


Poets that lasting marble seek,
Must carve in Latin or in Greek,
We write in sand, our language grows,
And like the tide, our work o'erflows.

-- Edmund Waller

INFORMATION TECHNOLOGY

MorphAdorner

NUPOS

Introduction to NUPOS and Morphology

This section details Martin Mueller's "NUPOS" part of speech tagset and makes explicit the structure of the tagset and other related morphology objects such as "spellings", "word classes", "lemmata", and "word parts".

As a convention, in this discussion, when we use the term "word", it means "a specific single occurrence of a word somewhere in a text." For the concept of a "word in general", we will use the terms "headword" and "lemma", which we'll define and discuss in detail later.

The full version of NUPOS can handle both Greek and English texts and part of speech tagging. Here we only describe the subset of NUPOS that deals with English. For more information, see Martin Mueller's fuller description.

Spellings

The first and most basic attribute of a word is its spelling. This may seem to be a simple concept, but especially for earlier texts from periods before spelling became regularized, it is useful to distinguish among several different meanings of the term "spelling". In NUPOS there are three different "spellings" for each word:

1. The "token spelling". This is the spelling of the word exactly as it appears in the original digital source for the text, including all capitalization and any typographical conventions that might be used in the source as markup for various purposes. For example, the original source for a text might contain a word token "common|lie", where the encoders used the vertical bar character "|" to mark up a soft hyphen at the end of a line. As another example, in some early printed texts, a "y" with a superscript "t" was used to represent the word "that". Such a word might be marked up as "y^t" in the source for such a text. As a final example, the token "@abper;fecit" might appear in the source for an early text. In this example "&abper;" is a symbol used in early typesetting as an abbreviation for "per" or "par".

The token spelling retains as much fidelity as possible with the original digital source. It will often contain various kinds of non-uniform markup, as used by the organizations that digitally encoded the texts. It may be of interest to some researchers, but most people will be more interested in the other two kinds of spellings.

The token spelling may be of importance in contexts where an application wishes to reproduce as much visual fidelity as possible with original printed texts when displaying the text to users.

2. The "standard original spelling". This is a version of the spelling with the typographical conventions normalized, and in most contexts is probably what

Standardize archaic English spellings to standard spelling with online example

- **Text Segmenter**
  Segment text into cohesive subtopics with online example
- **Verb Conjugator**
  Conjugate English verbs with online example
- **Word Tokenizer**
  Extract words from English text with online example

one thinks of when one uses the general term "the spelling of the word". It is usually identical with the token spelling, but not always. In the examples above, the three tokens become the following "standard original spellings":

```
common|lie --> commonlie
y^t --> that
@abper;fecit --> perfecit
```

3. The "standard modern spelling". This is the standard modern orthographic form of the original spelling. But the morphological form is not modernized. Thus a spelling like "lovyth" is regularized to "loveth". "loveth" is not, however, regularized to "loves", but is rather recognized as a standard archaic form. In the three examples above, the standard modern spellings are as follows:

```
common|lie --> commonlie --> commonly
y^t --> that --> that
@abper;fecit --> perfecit --> perfecit
```

Note that "perfecit" is a Latin word, and at no point is there an attempt made to translate foreign words into English.

For modern texts, the three spellings are nearly always identical. The main exceptions will be for words in XML texts split by decorator (soft) tags.

Word Parts

Words have spellings, as outlined above. We also want to enumerate and discuss in detail their other tagging attributes, such as word class, part of speech, and lemma. Before we can do this, however, we need to discuss a pesky complexity of texts - contractions.

Consider as an example the first word of *Hamlet*, "Who's". This is a single lexical word, and in this example all three spellings of the word are the same string "Who's".

In terms of the other attributes, however, this word is properly considered to be a lexical representation of the two separate words "who" and "is". Each part has its own word class, part of speech and lemma. In this particular example, it might also be possible to think of each part as having its own spelling or "sub-spelling", "who" and "'s", but in the general case it might be difficult to reasonably split up a spelling into its pieces, and the current version of NUPOS does not attempt to do this.

In NUPOS, this word "who's" is tagged as follows:

word part	major word class	word class	part of speech	lemma
1	wh-word	crq	q-crq	who (crq)
2	verb	va	vbz	be (va)

While we might wish that this complexity didn't exist or could be safely ignored, it can be important when analyzing texts. For example, consider the set of all words in Shakespeare which are instances of the auxiliary verb "be". In NUPOS, the first word of *Hamlet* is correctly included as a member of this set. It is also a member of the set of all words in Shakespeare which are instances of the wh-word "who".

As another example, consider the general notion of counting different kinds of words in Shakespeare. In NUPOS, the count of the total number of occurrences of the auxiliary verb "be" includes the first word of *Hamlet*, as it should, as does the count of the total number of occurrences of the wh-word "who". The first word of *Hamlet* is counted twice, once as "be" and once as "who". Consequently, the sum of the counts of the number of different kinds of words in *Hamlet* is equal to the number of word parts in *Hamlet*, not the number of words.

As a final example, consider an analysis of bigrams in Shakespeare. In NUPOS, the first word of *Hamlet* is considered to be an instance of the bigram "the lemma who (crq) followed by the lemma be (va)", as well as an instance of the bigram "word class crq followed by part of speech vaz".

In the general case, each word, while it usually only has one part, might have more than one part -- two parts in the case of most contractions, but at least conceivably perhaps even more than two parts. While it is words which possess spelling attributes, it is their parts which possess the other morphological attributes, and this is an important distinction to keep in mind.

In the normal case, when a word has only one part, we often use the simple

term "word" to refer to its unique part. For example, we say "this word is a verb", when to be precise what we are really saying is "the one and only part of this word is a verb."

Word Classes

In NUPOS, each word part has a "major word class" and a "word class". These concepts provide the coarsest ways to categorize words.

There are 17 major word classes, which should be self-explanatory:

Major word classes
adjective
adv/conj/pcl/prep
adverb
conjunction
determiner
foreign word
interjection
negative
noun
numeral
preposition
pronoun
punctuation
symbol
undetermined
verb
wh-word

Major word classes are subdivided into a slightly finer categorization by "word class". There are 34 word classes in NUPOS:

Name	Description	Major Class
acp	adverb/conjunction/particle/preposition	adv/conj/pcl/prep
an	adverb/noun	noun
av	adverb	adverb
cc	coordinating conjunction	conjunction
crq	wh-word	wh-word
cs	subordinating conjunction	conjunction
d	determiner	determiner
dt	article	determiner
fo	foreign	foreign word
fr	French	foreign word
ge	German	foreign word
gr	Greek	foreign word
it	Italian	foreign word
j	adjective	adjective
jn	adjective/noun	adjective
jp	proper adjective	adjective
la	Latin	foreign word
n	noun	noun
np	proper noun	noun
nu	numeral	numeral
pf	preposition "of"	preposition
pi	indefinite pronoun	pronoun
pn	personal pronoun	pronoun
po	possessive pronoun	pronoun
pp	preposition	preposition
pu	punctuation	punctuation
px	reflexive pronoun	pronoun
sy	symbol	symbol
uh	interjection	interjection

v	verb	verb
va	auxiliary verb	verb
vm	modal verb	verb
xx	negative	negative
zz	undetermined	undetermined

Each word class has a very short string which provides a name for the word class, and each word class belongs to one and only one of the major word classes.

For example, for the major word class "verb", there are three word classes "va" (auxiliary verb), "vm" (modal verb), and "v" (verb). So in NUPOS, there are three kinds of verbs.

Parts of Speech

NUPOS has a fine-grained part of speech tagset, much finer-grained than the word classes and major word classes. There are 241 total English parts of speech in the current version of NUPOS (not counting punctuation).

Each part of speech belongs to one and only one word class, so the part of speech tagset in NUPOS represents a subdivision of the word class tagset, in the same way that the word class tagset represents a subdivision of the major word class tagset.

To continue the example of verbs, in NUPOS each of the verb word classes contains a number of parts of speech:

```
word class va (auxiliary verb): 19 parts of speech
word class vm (modal verb): 14 parts of speech
word class v (verb): 27 parts of speech
```

Each part of speech, in addition to belonging to a word class, is also characterized by, and largely defined by, how it is used in various grammatical categories. These categories and their possible values should be mostly self-explanatory to those familiar with English grammar.

```
Syntax (used as): See below.
Tense: pres, past or empty (not applicable)
Mood: ppl, inf, impt or empty (not applicable)
Case: gen, obj, subj, or empty (not applicable)
Person: 1st, 2nd, 3rd, or empty (not applicable)
Number: sg, pl, or empty (not applicable).
Degree: comp, sup, or empty (not applicable).
Negative: no, nor, not, or empty (not applicable).
```

As an example, the NUPOS part of speech "vmd2" is used for modal verbs used in the second person singular past tense. It has the following attributes in addition to its name "vmd2":

```
word class = vm (modal verb)
syntax = vm
tense = past
mood = empty
case = empty
person = 2nd
number = sg
degree = empty
negative = empty
```

An example of this part of speech occurs in Act 5, Scene 1 of *Hamlet*, where Gertrude says "I hoped thou shouldst have been my Hamlet's wife;" In this passage, the word "shouldst" is tagged with the lemma "shall (vm)" and the part of speech "vmd2". By virtue of this tagging, we know all of the following facts about this word:

```
It is an instance of the headword "shall"
It is a verb.
It is a modal verb.
It has NUPOS part of speech "vmd2".
It is in the past tense.
It is in the second person.
It is singular.
```

In a full implementation of NUPOS, any of these attributes can be used as a criterion for searching, grouping, sorting, counting, and analysis. For example, a researcher might compare the use of past tense modal verbs by one author to their use by another author, or he might do a search where he finds all uses of second person singular verbs in the works of Chaucer. Or he might find all of the verbs used in Spenser and generate a report which counts up how many times each of them are used in the various possible

combinations of person and number.

The "syntax" attribute is used to specify how the part of speech is used. For example, the part of speech "av-j" is used for adjectives that are used as adverbs. The "syntax" attribute of this part of speech is "av". An example of this part of speech occurs in Act 1, Scene 1 of *Hamlet*, where Bernardo says "Long live the king!" The word "Long" in this passage in used as an adverb modifying the verb "live" and has the NUPOS part of speech "av-j". Contrast this with the word "long" in Act 3, Scene 1, where Hamlet says "That makes calamity of so long life;". In this passage, the word "long" is tagged with the part of speech "j", the part of speech for "normal" uses of adjectives. Both of the parts of speech "av-j" and "j" have the word class "j" and major word class "adjective", but "av-j" has the syntax attribute "av", while "j" has the syntax attribute "j".

Martin has also mentioned the possibility of more coarse-grained versions of NUPOS, finer grained than word classes but coarser than the full set of 220+ parts of speech. These intermediate levels of NUPOS may be useful for data mining and other kinds of analysis. We have not yet worked out the details of this idea.

Another distinctive feature of NUPOS is that it offers some ambiguous wordclasses, like 'jn' for words that hover between noun and adjective or 'an' for words that hover between noun and adverb (home, tomorrow).

All of the NUPOS parts of speech are displayed at the end of this appendix.

Lemmata

A lemma is a dictionary "headword" plus its word class.

For example, consider the verb "love" in Shakespeare. This lemma has the headword "love" and the word class "v". He uses this common lemma in 41 of his 42 works, a total of 1,135 times, in a variety of contexts with quite a few different parts of speech and spellings. For example, he uses it a total of 153 times with the part of speech "vvz", which is the NUPOS part of speech tag for verbs used in the third person singular in the present tense. 150 of these uses are spelled "loves", and three of them are spelled "loveth".

There is, of course, also a noun named "love". In NUPOS, there are two separate lemmata for the headword "love", one for the noun and one for the verb. In general, headwords like "love" are used to form NUPOS lemmata based on their word class, and the word class is listed along with the headword when naming the lemma. In our example, the NUPOS names for the two "love" lemmata are "love (n)" and "love (v)".

The set of all lemmata used in a work or collection of works is called the "lexicon" for the work or collection.

MorphAdorner

MorphAdorner reads source XML texts, locates sentence and word boundaries, and marks each word with five morphological tags -- the three spellings, the NUPOS part of speech, and the lemma headword. For contractions, MorphAdorner emits multiple parts of speech and headwords.

It's important to recall that MorphAdorner is more than just a part of speech tagger. It's also a spelling normalizer and a lemma tagger.

This tagging data emitted by MorphAdorner is sufficient to recover all of the information mentioned above for each word and word part, including the major word class, word class, part of speech category values, and lemma (headword plus major word class). Note that MorphAdorner only emits the lemma headword. The word class may be deduced from the part of speech.

Following the approach to contracted forms taken by NUPOS, Morphadorner treats contracted forms as a single token for two reasons.

1. The orthographic practice reflects an underlying linguistic reality that the tokenization should respect.
2. In Early Modern English (as in Shaw's orthographic reforms) contracted forms appear without apostrophes, as in 'noot' for 'knows not' or 'niltow' for 'wilt thou not'. It's not obvious how to split these forms. The situation is even less clear for dialectical forms.

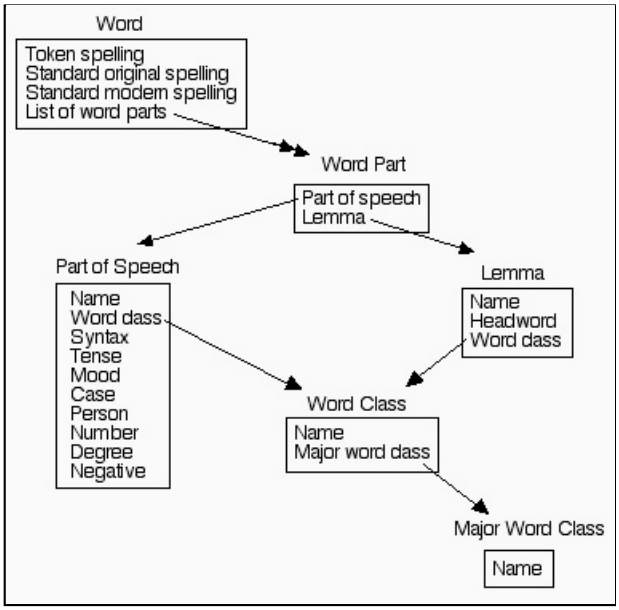
Contracted forms get two part of speech tags separated by a vertical bar, but with regard to forms like "don't", "cannot", "ain't", MorphAdorner analyzes the forms as the negative form of a verb and does not treat the form as a contraction. It uses the symbol 'x' to mark a negative part of speech tag.

Summary

NUPOS comprises the following objects, attributes, and relationships:

- Each word has three spellings: the token, standard original, and standard modern spellings.
- Each word has an ordered list of word parts, usually only one except for contractions.
- Each word part has a part of speech and a lemma.
- Each part of speech has a name, a word class, and values for the grammatical categories of syntax, tense, mood, case, person, number, degree, and negative.
- Each lemma has a name, a headword and a word class. The name of each lemma is formed from its headword and the name of its word class.
- Each word class has a name and a major word class.
- Each major word class has a name.
- In a full implementation of NUPOS, all of these objects and their attributes can be used as criteria for searching, grouping, sorting, counting, and analysis.

The following diagram is useful as a way of summarizing NUPOS. It's not a formal UML diagram, and the drawing has no particular implementation implications, other than as a way of summarizing some of the functionality that any particular full implementation of NUPOS must support. It's just an informal way of making a picture out of the objects, attributes, and relationships enumerated above and described and defined in detail in this note. The double-headed arrow is used to indicate the relationship "may have more than one of", while the single-headed arrow indicates "has one and only one of". The term "list of" in the one-to-many relationship between words and their parts indicates that the parts of a word are ordered -- there's a first one, then a second one, and so on. This is important for dealing with n-grams.



NUPOS for English

The following table lists all the non-punctuation parts of speech defined by NUPOS. The first column provides the NUPOS part of speech tag. The second column describes the tag. The third column offers an example the part of speech. The fourth column provides a rounded count of occurrences of the tag in the NUPOS training data expressed as parts per million. That shows how commonly a tag occurs in the MorphAdorner training data. The training data consists of about six million words drawn from the following texts:

- The complete works of Chaucer and Shakespeare
- Spenser's *Faerie Queene*
- North's translation of Plutarch's Lives
- Mary Wroth's *Urania*
- Jane Austen's *Emma*
- Dickens' *Bleak House* and *The Old Curiosity Shop*
- Emily Bronte's *Wuthering Heights*
- Thackeray's *Vanity Fair*
- Mrs. Gaskell's *Mary Barton*

- Frances Trollope's *Michael Armstrong*
- George Eliot's *Adam Bede*
- Scott's *Waverley*
- Harriet Beecher Stowe's *Uncle Tom's Cabin*
- Melville's *Moby Dick*

Examples are chosen for the most part from the training data.

Tag	Explanation	Example	Occurences per million words
a-acp	acp word as adverb	I have not seen him since	9,500
av	adverb	soon	37,500
av-an	noun-adverb as adverb	go home	750
av-c	comparative adverb	sooner, rather	500
avc-jn	comparative adj/noun as adverb	deeper	8
av-d	determiner/adverb as adverb	more slowly	2,000
av-dc	comparative determiner/adverb as adverb	can lesser hide his love	1,900
av-ds	superlative determiner as adverb	most often	900
av-dx	negative determiner as adverb	no more	600
av-j	adjective as adverb	quickly	15,500
av-jc	comparative adjective as adverb	he fared worse	850
av-jn	adj/noun as adverb	duly, right honourable	1,100
av-js	superlative adjective as adverb	in you it best lies	150
av-n1	noun as adverb	had been cannibally given	2
av-s	superlative adverb	soonest	14
avs-jn	superlative adj/noun as adverb	hee being the worthylest constant	0
av-vvg	present participle as adverb	lovingly	250
av-vvn	past participle as adverb	Stands Macbeth thus amazedly	85
av-x	negative adverb	never	1,300
c-acp	acp word as conjunction	since I last saw him	14,000
cc	coordinating conjunction	and, or	42,500
cc-acp	acp word as coordinating conjunction	but	6,500
c-crq	wh-word as conjunction	when she saw	6,500
ccx	negative conjunction	nor	1,200
crd	numeral	2, two, ii	5,700
cs	subordinating conjunction	if	6,500
cst	'that' as conjunction	I saw that it was hopeless	14,000
d	determiner	that man, much money	29,500
dc	comparative determiner	less money	850
dg	determiner in possessive use	the latter's	7
ds	superlative determiner	most money	450
dt	article	a man, the man	7,000
dx	negative determiner as adverb	no money	2,500
fw-fr	French word	monsieur	500
fw-ge	German word	Herr	15
fw-gr	Greek word	kurios	15
fw-it	Italian word	signor	10
fw-la	Latin word	dominus	400
fw-mi	word in unspecified other language	n/a	50
j	adjective	beautiful	49,500
j-av	adverb as adjective	the then king	1
jc	comparative adjective	handsomer	1,500
jc-jn	comparative adj/noun	yet she much whiter	70

jc-vvg	present participles as comparative adjective	for what pleasinger then varietie, or sweeter then flatterie?	1
jc-vvn	past participle as comparative adjective	shall find curster than she	1
j-jn	adjective-noun	the sky is blue	7,000
jp	proper adjective	Athenian philosopher	800
js	superlative adjective	finest clothes	1,500
js-jn	superlative adj/noun	reddest hue	200
js-vvg	present participle as superlative adjective	the lyingest knave in Christendom	2
js-vvn	past participle as superlative adjective	deformed'st creature	3
j-vvg	present participle as adjective	loving lord	2,000
j-vvn	past participle as adjective	changed circumstances	2,500
n1	singular, noun	child	14,000
n1-an	noun-adverb as singular noun	my home	250
n1-j	adjective as singular noun	a good	4
n2	plural noun	children	35,000
n2-acp	acp word as plural noun	and many such-like "As'es" of great charge	1
n2-an	noun-adverb as plural noun	all our yesterdays	9
n2-av	adverb as plural noun	and are etcecteras no things	1
n2-dx	determiner/adverb negative as plural noun	yeas and honest kerysey noes	0
n2-j	adjective as plural noun	give me particulars	200
n2-jn	adj/noun as plural noun	the subjects of his substitute	600
n2-vdg	present participle as plural noun, 'do'	doings	50
n2-vhg	present participle as plural noun, 'have'	my present havings	1
n2-vvg	present participle as plural noun	the desperate languishings	200
n2-vvn	past participle as plural noun	there was no necessity of a Letter of Slains for Mutilation	0
ng1	singular possessive, noun	child's	2,500
ng1-an	noun-adverb in singular possessive use	Tomorrow's vengeance	6
ng1-j	adjective as possessive noun	the Eternal's wrath	1
ng1-jn	adj/noun as possessive noun	our sovereign's fall	60
ng1-vvn	past participle as possessive noun	the late lamented's house	0
ng2	plural possessive, noun	children's	350
ng2-jn	adj/noun as plural possessive noun	mortals' chiefest enemy	50
n-jn	adj/noun as noun	a deep blue	2,300
njp	proper adjective as noun	a Roman	130
njp2	proper adjective as plural noun	The Romans	1,300
njpg1	proper adjective as possessive noun	The Roman's courage	8
njpg2	proper adjective as plural possessive noun	The Romans' courage	20
np1	singular, proper noun	Paul	27,500
np2	plural, proper noun	The Nevils are thy subjects	350
npg1	singular possessive, proper noun	Paul's letter	2,600
npg2	plural possessive, proper noun	will take the Nevils' part	6
np-n1	singular noun as proper noun	at the Porpentine	260

np-n2	plural noun as proper noun	such Brooks are welcome to me	2
np-ng1	singular possessive noun as proper noun	and through Wall's chink	20
n-vdg	present participle as noun, 'do'	my doing	20
n-vhg	present participle as noun, 'have'	my having	0
n-vvg	present participle as noun	the running of the deer	1,500
n-vvn	past participle as noun	the departed	50
ord	ordinal number	fourth	2,500
p-acp	acp word as preposition	to my brother	57,000
pc-acp	acp word as particle	to do	19,000
pi	singular, indefinite pronoun	one, something	2,200
pi2	plural, indefinite pronoun	from wicked ones	50
pi2x	plural, indefinite pronoun	To hear my nothings monstered	2
pig	singular possessive, indefinite pronoun	the pairings of one's nail	35
pigx	possessive case, indefinite pronoun	nobody's	2
pix	indefinite pronoun	none, nothing	1,300
pn22	2nd person, personal pronoun	you	9,000
pn31	3rd singular, personal pronoun	it	10,500
png11	1st singular possessive, personal pronoun	a book of mine	220
png12	1st plural possessive, personal pronoun	this land of ours	35
png21	2nd singular possessive, personal pronoun	this is thine	3
png22	2nd person, possessive, personal pronoun	this is yours	100
png31	3rd singular possessive, personal pronoun	a cousin of his	200
png32	3rd plural possessive, personal pronoun	this is theirs	30
pno11	1st singular objective, personal pronoun	me	5,000
pno12	1st plural objective, personal pronoun	us	1,100
pno21	2nd singular objective, personal pronoun	thee	1,200
pno31	3rd singular objective, personal pronoun	him, her	12,000
pno32	3rd plural objective, personal pronoun	them	4,700
pns11	1st singular subjective, personal pronoun	I	14,500
pns12	1st plural subjective, personal pronoun	we	2,200
pns21	2nd singular subjective, personal pronoun	thou	2,000
pns31	3rd singular subjective, personal pronoun	he, she	21,000
pns32	3rd plural objective, personal pronoun	they	5,600
po11	1st singular, possessive pronoun	my	6,700
po12	1st plural, possessive pronoun	our	1,400
po21	2nd singular, possessive pronoun	thy	1,650
po22	2nd person possessive pronoun	your	3,000
po31	3rd singular, possessive pronoun	its, her, his	19,000
po32	3rd plural, possessive pronoun	their	3,800

pp	preposition	in	23,000
pp-f	preposition 'of'	of	29,000
px11	1st singular reflexive pronoun	myself	350
px12	1st plural reflexive pronoun	ourselves	55
px21	2nd singular reflexive pronoun	thyself, yourself	250
px22	2nd plural reflexive pronoun	yourselves	30
px31	3rd singular reflexive pronoun	herself, himself, itself	1,300
px32	3rd plural reflexive pronoun	themselves	220
pxg21	2nd singular possessive, reflexive pronoun	yourself's remembrance	1
q-crq	interrogative use, wh-word	Who? What? How?	3,000
r-crq	relative use, wh-word	the girl who ran	10,000
sy	alphabetical or other symbol	A, @	50
uh	interjection	oh!	3,000
uh-av	adverb as interjection	Well!	300
uh-crq	wh-word as interjection	Why, there were but four	500
uh-dx	negative interjection	No!	500
uh-j	adjective as interjection	Grumio, mum!	7
uh-jn	adjective/noun as interjection	And welcome, Somerset	30
uh-n	noun as interjection	Soldiers, adieu!	200
uh-v	verb as interjection	My gracious silence, hail	90
vb2	2nd singular present of 'be'	thou art	300
vb2-imp	2nd plural present imperative, 'be'	Beth pacient	10
vb2x	2nd singular present, 'be'	thow nart yit blisful	2
vbb	present tense, 'be'	are, be	3,300
vbbx	present tense negative, 'be'	aren't, ain't, beant	60
vbd	past tense, 'be'	was, were	14,000
vbd2	2nd singular past of 'be'	thou wast, thou wert	50
vbd2x	2nd singular past, 'be'	weren't	0
vbdp	plural past tense, 'be'	whose yuorie shoulders weren couered all	30
vbdx	past tense negative, 'be'	wasn't, weren't	75
vbg	present participle, 'be'	being	1,300
vbi	infinitive, 'be'	be	5,600
vbm	1st singular, 'be'	am	1,200
vbmx	1st singular negative, 'be'	I nam nat lief to gabbe	3
vbn	past participle, 'be'	been	1,800
vbp	plural present, 'be'	Thise arn the wordes	260
vbz	3rd singular present, 'be'	is	6,900
vbzx	3rd singular present negative, 'be'	isn't	100
vd2	2nd singular present of 'do'	dost	150
vd2-imp	2nd plural present imperative, 'do'	Dooth digne fruyt of Penitence	6
vd2x	2nd singular present negative, 'do'	thee dostna know the pints of a woman	2
vdb	present tense, 'do'	do	1,600
vdbx	present tense negative, 'do'	don't	500
vdd	past tense, 'do'	did	3,100
vdd2	2nd singular past of 'do'	didst	55
vdd2x	2nd singular past negative, verb	Why, thee thought'st Hetty war a ghost, didstna? 0.20	
vddp	plural past tense, 'do'	on Job , whom that we diden wo	3
vddx	past tense negative, 'do'	didn't	90
vdg	present participle, 'do'	doing	110
vdi	infinitive, 'do'	to do	1,000
vdn	past participle, 'do'	done	700
vdp	plural present, 'do'	As freendes doon whan they been met	30

vdz	3rd singular present, 'do'	does	800
vdzx	3rd singular present negative, 'do'	doesn't	20
vh2	2nd singular present of 'have'	thou hast	250
vh2-imp	2nd plural present imperative, 'have'	O haveth of my deth pitee!	1
vh2x	2nd singular present negative, 'have'	hastna	0
vhb	present tense, 'have'	have	2,500
vhbx	present tense negative, 'have'	haven't	30
vhd	past tense, 'have'	had	6,000
vhd2	2nd singular past of 'have'	thou hadst	35
vhdp	plural past tense, 'have'	Of folkes that hadden grete fames	10
vhdx	past tense negative, 'have'	hadn't	20
vhg	present participle, 'have'	having	730
vhi	infinitive, 'have'	to have	2,400
vhn	past participle, 'have'	had	220
vhp	plural present, 'have'	They han of us no jurisdiccioun,	120
vhz	3rd singular present, 'have'	has, hath	1,700
vhzx	3rd singular present negative, 'have'	Ther loveth noon, that she nath why to pleyne.	11
vm2	2nd singular present of modal verb	wilt thou	360
vm2x	2nd singular present negative, modal verg	O deth, allas, why nyltow do me deye	4
vmb	present tense, modal verb	can, may, shall, will	8,300
vmb1	1st singular present, modal verb	Chill not let go, zir, without vurther 'cagion	3
vmbx	present tense negative, modal verb	cannot; won't; I nyl nat lye	700
vmd	past tense, modal verb	could, might, should, would	8,300
vmd2	2nd singular past of modal verb	couldst, shouldst, wouldst; how gret scorn woldestow han	120
vmd2x	2nd singular present, modal verb	Why noldest thow han writen of Alceste	5
vmdp	plural past tense, modal verb	tho thinges ne scholden nat han ben doon	30
vmdx	past negative, modal verb	couldn't; She nolde do that vileynye or synne	160
vmi	infinitive, modal verb	Criseyde shal nought konne knowen me.	5
vmn	past participle, modal verb	I had oones or twyes ycould	2
vmp	plural present tense, modal verg	and how ye schullen usen hem	25
vv2	2nd singular present of verb	thou knowest	480
vv2-imp	2nd present imperative, verb	For, sire and dame, trusteth me right weel,	80
vv2x	2nd singular present negative, verb	"Yee!" seyde he, "thow nost what thow menest;	1
vvb	present tense, verg	they live	17,000
vvbx	present tense negative, verb	What shall I don? For certes, I not how	30
vvd	past tense, verb	knew	33,000
vvd2	2nd singular past of verb	knewest	75
vvd2x	2nd singular past negative, verb	thou seidest that thou nystist nat	0
vvdp	past plural, verb	They neuer strouen to be chiefe	80
vvdx	past tense negative, verb	she caredna to gang into the stable	10
vvg	present participle, verb	knowing	13,700

vvi	infinitive, verb	to know	36,000
vvn	past participle, verb	known	26,200
vvp	plural present, verb	Those faytours little regarden their charge	330
vvz	3rd singular preseent, verb	knows	7,200
vvzx	3rd singular present negative, verb	She caresna for Seth.	1
xx	negative	not	7,800
zz	unknown or unparsable token	n/a	200