Verb Analysis

Verb Tense, Voice, Modality, and Mood....1
Tense....1
Voice....2
Passive voice and agency....2
Modality....2
Mood....3
Verb mood are not currently computed by Stanford CoreNLP....3
References....3

There is evidence that men's and women's writing show different frequencies of nouns and verbs. Men use more nouns; women more verbs (verbs being more relational than nouns).

Do your documents show a differential use of verbs?

Does the use of verbs vary along a text, if you plot verbs by sentence index?

Verb Tense, Voice, Modality, and Mood

Verbs are perhaps the most important words in sentences. Four different aspects of verbs are particularly important: Tense, Mood, Voice, and Modality. Computational linguistics parsers, such as the Stanford CoreNLP or NLTK, make it easy to tease out automatically these aspects of language, as they are colntained in the POSTAG (Part of Speech Tags) and DEPREL (Dependency Relation) fields of the CoNLL table.

There is evidence of a differential use of nouns versus verbs and of verb tenses. Females use more verbs (especially auxiliary verbs), while men use more nouns, especially concrete nouns. Paradoxically, the analysis of auxiliary verbs shows that *older people* use more future tense and less past tense the older they get (see Chung and Pennebaker 2007).

Tense

Tense refers to a measure of **time:** past, present, and future, although not all actions fit precisely into those three categories. Sometimes actions start in the past and continue into the present, and they may go on into the future (gerundive).

In the CoNLL table verb tenses are found in these POSTAG (Part of Speech Tags): VB, VBD, VBG, VBN, VBP, VBZ.

How does a text use verb tense? Is there a consistent use of time, i.e., does a text start in one time and stays in it or does it jump back and forth between tenses?

There is evidence of a differential use of verb tenses. Paradoxically, older people use the future more frequently thhan they use the present or past ((see Chung and Pennebaker 2007).

Voice

A verb voice tells us whether the subject of a sentence is the actor (agent or sender) or is acted upon (patient or reciver). While agency (who does something) is *always* present in a sentence constructed with an active verb, with passive verbs agency can be denied. For example, in the sentence "the police atacked the demonstrator" the syntactical subject and semantic role coincide (police); in the passive form of the sentence, "the demonstrators were attacked by the police", subject (the demonstrators) and the semantic role of agent (the police) are different. In a passive voice, agency can be alltogether omitted and still have a syntactically correct sentence ("the demonstrators were atacked").

Active voice example: John ate the apple.
Passive voice example: The apple was eaten by John.

In the CoNLL table verb voice is found in a combination of DEPREL (Dependency Relation) and POSTAG (Part of Speech Tags) values. All verbs with DEPREL value "auxpass" (auxialiary passive) *and* all verbs with POSTAG value "VBN" (past participle) are counted as passive. All verbs with any POSTAG values starting with VB* and not already counted as passive are counted as active.

Passive voice and agency

Together with nominalization (i.e., the turning of a verb into a noun, e.g., the lynching, the attack, the rape), passive voices can be taken as a sign of **denial of agency**. In an active sentence, syntactic subject and semantic role of agent overlap and the agent is marked ("the police charged the demonstrators", the police is bot subject and agent); in a passive voice ("the demonstrators were charged by the police"), the subject is now the demonstrators and the agent is unmarked. Furthermore, the agent can be omittted altogether in a passive sentence and still produce a gramatically correct sentence ("the demonstrators were charged"). You may also run noun analysis on the use of passive nouns. Which nouns, particularly social actors, are used in the passive form?

See Franzosi et al. (2012).

Modality

Modality refers to verbs expressing ability possibility permission and obligation. Examples of modal verbs in English are: can/could, may/might, must/need, will/would and shall/should/ought. Modality is about a speaker's or a writer's attitude towards the world. A speaker or writer can express certainty, possibility, willingness, obligation, necessity and ability by using modal words and expressions.

In the CoNLL table verb modality is found in the POSTAG (Part of Speech Tags) field with value: MD. We look at these FORM values to further classify modality in specific categories:

Likelihood/possibility: FORM = can/could/may/might, will/'ll/would/'d
Ability: FORM = can/could
Permission: FORM = can/could/may/might
Obligation: FORM = will/would, must/need, shall/should/ought

Some FORM values (e.g., can/could) are counted as many times as they appear in specific modality classes (i.e., for can/could in Likelihood/possibility, Ability, Permission). Only context can provide clues as to the correct type of classification.

Mood

Verbs have several moods:

Indicative: used for statements of fact and for asking questions
Subjunctive: makes a statement contrary to fact (e.g., "If I were twenty, I would definitely go to college." Unfortunately, the speaker happens to be sixty, so this is contrary to fact).
Imperative: expresses a request or command
Conditional: used for statements that are dependent on each other; one statement is true only if another statement is true.

Verb mood are not currently computed by Stanford CoreNLP

It should be noted that CoreNLP sometimes parses sentences wrongly which may lead to wrong counts. For instance, the sentence found in the sample database newspaper article, Atlanta Constitution_2-10-1888_2, "The negro was carried in to woods, and it is supposed he was hung or burned." Is parsed in the following way:

1	The	the	DT	O	2	det	217	12
2	negro	negro	NN	O	4	nsubjpass	218	12
3	was	be	VBD	O	4	auxpass	219	12
4	carried	carry	VBN	O	0	ROOT	220	12
5	in	in	IN	O	7	case	221	12
6	to	to	TO	O	7	case	222	12
7	woods	wood	NNS	O	4	nmod	223	12
8	,	,	,	O	4	punct	224	12
9	and	and	CC	O	4	cc	225	12
10	it	it	PRP	O	12	nsubjpass	226	12
11	is	be	VBZ	O	12	auxpass	227	12
12	supposed	suppose	VBN	O	4	conj	228	12
13	he	he	PRP	O	15	nsubj	229	12
14	was	be	VBD	O	15	cop	230	12
15	hung	hung	JJ	O	12	ccomp	231	12
16	or	or	CC	O	12	cc	232	12
17	burned	burn	VBN	O	12	conj	233	12
18	.	.	.	O	4	punct	234	12

The word "hung" is parsed with POSTAG value "JJ", i.e., adjective. Although, indeed "hung" can be an adjective (e.g., "a hung jury"), in this case it is a verb (past principle) like the word "burned", correctly parsed. This is part of that 3% or so parsing errors of CoreNLP.

References

Argamon, Shlomo, Moshe Koppel, Jonathan Fine, and Anat Rachel Shimoni. 2003a. "Gender, Genre, and Writing Style in Formal Written Texts," *Text*, Vol. 23, No. 3, pp. 321–346.

Chung, Cindy and James Pennebaker. 2007. "The Psychological Functions of Function Words." In: pp. 343-359, Klaus Fiedler (Ed.), *Social Communication*, New York: Psychology Press.

Franzosi, Roberto, Gianluca De Fazio, and Stefania Vicari. 2012. "Ways of Measuring Agency and Action: An Application of Quantitative Narrative Analysis to Lynchings in Georgia (1875-1930)." In: pp. 1-41, Tim Liao (ed.), *Sociological Methodology*, Vol. 42.