Clausal Analysis

Clausal analysis: Definition

Clausal analysis involves the division of sentences into their constituent clauses: for example, the analysis of the sentence "When the bell rang, all the children ran out of the classroom." "all the children ran out of the classroom" is the main clause, and "When the bell rang" is a subordinate adverbial clause of time. This type of analysis was once standard practice in schools, based on the study of Latin. But from the 1960's the practice was slowly abandoned in the English-speaking world and students have generally had little organized instruction in sentence forms. Natural Language Processing (NLP) parsers provide automatic ways of clausal analysis.

Stanford CoreNLP parser

Clausal analysis is based on the CoNLL table produced by the Stanford PCFG Parser (http://nlp.stanford.edu/software/lex-parser.shtml).

The Stanford CoreNLP neural network parser does NOT produce clausal tags (only the PCFG parser).

The tags used by the Stanford CoreNLP parser are based on the Penn Treebank Project (https://web.archive.org/web/19970614160127/http://www.cis.upenn.edu/~treebank/)

Clause Level Tags

S	Sentence – simple declarative clause, i.e. one that is not introduced by a (possible empty) subordinating conjunction or a wh-word and that does not exhibit subject-verb inversion.
SBAR	Clause introduced by a (possibly empty) subordinating conjunction.
SBARQ	Direct question introduced by a wh-word or a wh-phrase. Indirect questions and relative clauses should be bracketed as SBAR, not SBARQ.
SINV	Inverted declarative sentence, i.e. one in which the subject follows the tensed verb or modal.
SQ	Inverted yes/no question, or main clause of a wh-question, following the wh-phrase in SBARQ.

Phrase Level Tags

ADJP	Adjective Phrase.

ADVP	Adverb Phrase.
CONJP	Conjunction Phrase.
FRAG	Fragment.
INTJ	Interjection. Corresponds approximately to the part-of-speech tag UH.
LST	List marker. Includes surrounding punctuation.
NAC	Not a Constituent; used to show the scope of certain prenominal modifiers within an NP.
NP	Noun Phrase.
NX	Used within certain complex NPs to mark the head of the NP. Corresponds very roughly to N-bar level but used quite differently.
PP	Prepositional Phrase.
PRN	Parenthetical.
PRT	Particle. Category for words that should be tagged RP.
QP	Quantifier Phrase (i.e. complex measure/amount phrase); used within NP.
RRC	Reduced Relative Clause.
UCP	Unlike Coordinated Phrase.
VP	Verb Phrase.
WHADJP	Wh-adjective Phrase. Adjectival phrase containing a wh-adverb, as in how hot.
WHAVP	Wh-adverb Phrase. Introduces a clause with an NP gap. May be null (containing the 0 complementizer) or lexical, containing a wh-adverb such as how or why.
WHNP	Wh-noun Phrase. Introduces a clause with an NP gap. May be null (containing the 0 complementizer) or lexical, containing some wh-word, e.g. who, which book, whose daughter, none of which, or how many leopards.
WHPP	Wh-prepositional Phrase. Prepositional phrase containing a wh-noun phrase (such as of which or by whose authority) that either introduces a PP gap or is contained by a WHNP.
X	Unknown, uncertain, or unbracketable. X is often used for bracketing typos and in bracketing the...the-constructions.

What clausal analysis says about style

Sentences can be constructed in very different ways, with a mixture of the different types of clauses, of different types of combinations of phrase types. As John O'Hayre writes in his *Goobledygook Has Gotta Go* (1966):

> Inversion well handled makes for true and interesting variety. Winston Churchill, a master of the long sentence, was also a master of the inverted sentence; see how effectively he uses inversion in this sentence from *The Birth of Britain*: "'You will beat them,' he said, and-marking the town of Preston with his thumbnail on the map – 'you will beat them there!' And on November 13, beaten there they were."

Clausal measures (e.g., SINV for inversion) and plots in the NLP Suite provide ways of tapping an author's style.
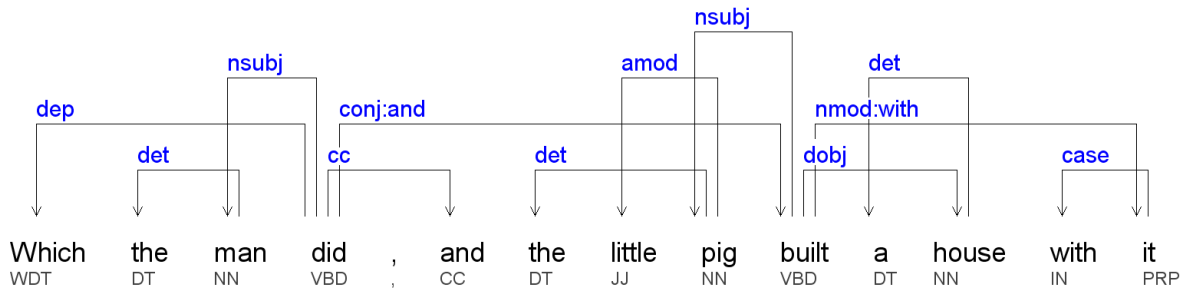
See also the TIPS file on style.

Other useful tools

Sentence complexity

You can compute measures of sentence complexity using the java script Sentence_Complexity.jar.

Visualizing the parse tree

Do try visualizing the parse tree using the java script dependenSee.jar.

dep nsubj conj:and amod nsubj det

det cc det nmod:with

dobj case

| Which | the | man | did | , | and | the | little | pig | built | a | house | with | it | . |
| WDT | DT | NN | VBD | , | CC | DT | JJ | NN | VBD | DT | NN | IN | PRP | . |

Text readability

The text readability script provides several measures of readability, in terms of how many years of formal education it takes to understand a text, starting from grade 1 (of the American education system).

Unusual words

Unusual words (and not just because a word is misspelled) do not contribute to sentence complexity at the syntactical level but they do contribute to text readability at the semantic level. Run the NLTK algorithm to get a list of unusual words in your documents.

n-grams and hapax legomena

n-grams and words that only occur once in a text (called hapax legomena, singular legomenon, often abbreviated to hapax) often provide clues to an author's use of language.

References

O'Hayre, John. 1966. *Goobledygook Has Gotta Go*. Washington, D.C.: U.S. Government Printing Office No. 0-206-141.