

# Comprehensive Guide to Croft's Flat Syntax CE Labels

## PHRASAL CE LABELS

\*Labels for elements within phrases (word-level components)\*

Label	Full Name	Definition	Function	Examples from Paper	Additional Examples
<b>Head</b>	Head	The █ element of a phrase that determines its referential or semantic properties; can stand alone as a minimal phrase	Denotes the referent or core meaning; syntactically required element	<ul style="list-style-type: none"> <li>"brothers" in "The two <b>brothers</b>"</li> <li>"lose" in "might <b>lose</b>"</li> <li>"competition" in "the <b>competition</b>"</li> </ul>	<ul style="list-style-type: none"> <li>"dog" in "the big <b>dog</b>"</li> <li>"quickly" in "very <b>quickly</b>"</li> <li>Single-word phrases are annotated as Head (§3.1)</li> </ul>
<b>Mod</b>	Modifier	Elements that specify, describe, or delimit the head referent; includes articles, demonstratives, numerals, quantifiers, adjectives	Restricts or elaborates the reference/meaning of the head	<ul style="list-style-type: none"> <li>"The two" in "<b>The two brothers</b>"</li> <li>"the" in "<b>the game</b>"</li> <li>"a rare but aggressive" in "<b>a rare but aggressive cancer</b>" (Table 3)</li> </ul>	<ul style="list-style-type: none"> <li>"big" in "the <b>big dog</b>"</li> <li>"this" in "<b>this book</b>"</li> <li>"three" in "<b>three cats</b>"</li> <li>Subtypes not distinguished: Art, Dem, Num, Adj all labeled Mod (§3.1)</li> </ul>
<b>Adm</b>	Admodifier	Modifiers of modifiers; degree words, intensifiers, downtoners, hedging elements that modify adjectives, adverbs, or other modifiers	Grades or qualifies the property expressed by a modifier	<ul style="list-style-type: none"> <li>"very" in "a <b>very</b> large tree" (Table 2)</li> <li>"almost" in "<b>almost</b> fifty" (§2.1)</li> </ul>	<ul style="list-style-type: none"> <li>"extremely" in "<b>extremely</b> hot"</li> <li>"quite" in "<b>quite</b> interesting"</li> <li>"somewhat" in "<b>somewhat</b> difficult"</li> <li>No separate attributive phrase level created (§2.1)</li> </ul>
<b>Adp</b>	Adposition	Prepositions or postpositions; mark spatial, temporal, or abstract relations between phrases	Strategy for expressing modification or argument relations	<ul style="list-style-type: none"> <li>"on" in "<b>on</b> top of" (before becoming MWE)</li> <li>"on" in "fell <b>on</b> (our) house" (Table 4)</li> </ul>	<ul style="list-style-type: none"> <li>English: <b>in</b>, <b>at</b>, <b>from</b>, <b>to</b></li> <li>Turkish: ev-<b>de</b> 'house-at' (postposition)</li> <li>Found in phrases but indicate clause-level CEs (§3.2, Table 7)</li> </ul>
<b>Lnk</b>	Linker	Grammatical elements connecting modifiers to heads; linking particles or morphemes	Strategy for marking modification relationships	<ul style="list-style-type: none"> <li>Spanish "de" in possessive constructions</li> <li>Tagalog linker particles (implied from typological context)</li> </ul>	<ul style="list-style-type: none"> <li>Mandarin "de" (█): hóng <b>de</b> shí 'red <b>DE</b> book'</li> <li>Malay "yang": rumah <b>yang</b> besar 'house <b>THAT</b> big'</li> <li>Found in both phrasal and clausal contexts (Tables 6-7)</li> </ul>
<b>Cif</b>	Classifier	Numerical or noun classifiers categorizing referents by shape, animacy, or other properties	Strategy for expressing quantification or categorization	<ul style="list-style-type: none"> <li>Not exemplified in English examples</li> <li>Referenced in typological discussion (§3.1)</li> </ul>	<ul style="list-style-type: none"> <li>Mandarin: sì n bì n shí 'three <b>CLF</b> book'</li> <li>Thai: rót sàam khan 'car three <b>CLF</b>'</li> <li>Japanese: hon ni-satsu 'book two-<b>CLF</b>'</li> </ul>
<b>Idx</b>	Index	Agreement markers; person/number/gender indexes that cross-reference referents (when realized as separate words)	Strategy for marking grammatical relations through indexation	<ul style="list-style-type: none"> <li>Rare in English</li> <li>Implied in discussion of "indexation" throughout §3.2</li> </ul>	<ul style="list-style-type: none"> <li>Swahili: <b>wa</b>-toto <b>wa</b>-kubwa 'children big' (both have CL2 prefix)</li> <li>Usually bound morphemes, but labeled when orthographically separate (§3.1)</li> </ul>

<b>Conj</b>	Conjunction (Phrasal)	Coordinators linking modifiers or phrases within a larger phrase structure	Strategy for coordinating elements at phrase level	• "and" in "the adults <b>and</b> the kids" (Table 3) • "but" in "a rare <b>but</b> aggressive cancer" (Table 3)	• "or" in "tea <b>or</b> coffee" • "both...and" in " <b>both</b> the priest <b>and</b> the shopkeeper" (ex. 2) • Comma in lists: "red, white, <b>and</b> blue"
-------------	-----------------------	--	--	--	--

---

## CLAUSAL CE LABELS

\*Labels for elements within clauses (phrase-level and some word-level components)\*

Label	Full Name	Definition	Function	Examples from Paper	Additional Examples
<b>Pred</b>	Predicate	The primary predicate of the clause; the main verb or predicate element expressing the event/state	Core semantic element denoting the situation	• "lose" in "might + <b>lose</b> " (Table 1) • "wrecked" in "also + <b>wrecked</b> + your + car" (Table 4) • "fell" in "{that + fell + on {our} house}" (Table 4)	• "runs" in "The dog <b>runs</b> " • "is" in "Ivan <b>is</b> the best dancer" (ex. 5b) • In complex predicates, the primary predicate only (§3.1)
<b>Arg</b>	Argument	Referring phrases denoting participants in the event; includes subjects, objects, and obliques—NOT differentiated into subtypes	Expresses participants/entities involved in the situation	• "The two brothers" in " <b>The two brothers</b> + might + lose" (Table 1) • "the game" in "lose + <b>the game</b> " (Table 1) • "the competition" in "exit + <b>the competition</b> " (Table 1)	• "She" in " <b>She ate</b> " • "the book" in "read <b>the book</b> " • "to school" in "walked <b>to school</b> " • Broad definition includes "adjuncts" (§3.1)
<b>CPP</b>	Complex Predicate Part	Non-primary parts of complex predicates; auxiliaries, manner adverbs, tense-aspect-modality elements occurring separately from main predicate	Elements contributing to complex predication without being the primary predicate	• "might" in " <b>might</b> + lose" (Table 1) • "also" in " <b>also</b> + wrecked" (Table 4) • "had" in " <b>had</b> + died" (implied from text)	• "can" in " <b>can</b> run" • "will" in " <b>will</b> go" • "been" in "has <b>been</b> eating" • "carefully" in "drives <b>carefully</b> " (adverb) • Not grouped into predicate phrases due to non-contiguity (§3.1)
<b>Gen</b>	Genitive Phrase	Possessive or genitive phrases modifying another nominal referent	Expresses possessive relationships between referents	• "our" in "on {our} house" (Table 4) • "your" in "wrecked + <b>your</b> + car" (Table 4) • "the dog's" in " <b>the dog's</b> + toy" (ex. 15a)	• "Mary's" in " <b>Mary's</b> book" • "of the table" in "the surface <b>of the table</b> " (§4.2.2) • May be single words (pronouns) or full phrases (§3.2)
<b>FPM</b>	Flagged Phrase Modifier	Phrases marked with adpositions or case that modify nominal referents; typically spatial, temporal, or other semantic relations	Expresses location, time, or other relations between referents	• Implied in discussion of "adpositional and case-marked noun modifiers" (§3.1) • Contrasted with Gen in Table 5	• "on the table" in "the book <b>on the table</b> " • "from Boston" in "the train <b>from Boston</b> " • Turkish: masa-da-ki kitap 'table-LOC-REL book' ("book on the table") • Grouped with Gen as "flags" in typology (§3.2, Table 7)

<b>Conj</b>	Conjunction (Clausal)	Coordinators or linking elements joining clauses	Strategy for connecting clauses	• "and" in "lose + the game # <b>and</b> + exit" (Table 1, ex. 5c) • "and" in "buy <b>and</b> eat" (ex. 5c)	• "but" in "I tried <b>but</b> I failed" • "or" in "Stay <b>or</b> go" • "so" in "exhausted and ( <b>so</b> ) she went to bed" (ex. 22b') • Found at both phrase and clause levels (Tables 5, 6, 7)
-------------	-----------------------	--	---------------------------------	---	---

## SENTENTIAL CE LABELS

\*Labels for elements within sentences (clause-level components)\*

Label	Full Name	Definition	Function	Examples from Paper	Additional Examples
<b>Main</b>	Main Clause	Clauses expressing pragmatically asserted events/states; independent clauses that could stand alone	Primary assertion in the utterance	• Both clauses in coordinate structure: "lose the game # <b>and</b> + exit the competition" (both labeled Main in Table 1)	• "I ate" in "I <b>ate</b> and left" • "She left" in " <b>She left</b> " • May occur with subordinate clauses: " <b>I left</b> when he arrived" • Distinguished by assertion, not morphosyntax alone (§4.1)
<b>Adv</b>	Adverbial Clause	Subordinate clauses expressing temporal, causal, conditional, or other circumstantial relations to main event	Provides circumstantial/temporal/causal background	• "after + I + ate" in " <b>after + I + ate</b> # I + left" (ex. 14a) • "If you do that" in " <b>If you do that</b> , the terrorists have won" (ex. 22d)	• "when he arrived" in "I left <b>when he arrived</b> " • "because she was exhausted" in "She went to bed <b>because she was exhausted</b> " (ex. 22b) • "Although John had no money" in " <b>Although John had no money</b> , he went into this restaurant" (ex. 22e)
<b>Rel</b>	Relative Clause	Clauses functioning as modifiers of nominal referents; attributive clauses	Restricts or describes the referent of a noun phrase	• "that + fell + on {our} house" in "The tree <b>{that + fell + on {our} house}</b> had + died" (Table 4) • "that + Gollum + had + lost" in "the ring # <b>that + Gollum + had + lost</b> " (ex. 16)	• "who is tall" in "the man <b>who is tall</b> " • "that I bought" in "the bicycle <b>that I bought</b> " (§3.2) • May interrupt matrix clause (center-embedding) or follow head (§4.2.1)
<b>Comp</b>	Complement Clause	Clauses functioning as arguments of predicates; complete propositions filling argument slots	Serves as subject, object, or complement of matrix predicate	• "that + he + resigned" in "It + isn't + surprising # <b>that + he + resigned</b> " (ex. 35b) • "to believe the girl" in "wants to <b>believe the girl</b> " (ex. 6, AMR example)	• "that I bought a bicycle" in "I told her <b>that I bought a bicycle</b> " (§3.2) • "to leave" in "She wants to <b>leave</b> " • May occur with utterance, attitude, perception, modal verbs (§4.2.2)

Dtch	Detached Phrase	Phrases that are pragmatically foregrounded as topics or foci; not integrated into clause argument structure	Marks pragmatic status (topic/focus) separate from clausal structure	<ul style="list-style-type: none"> <li>Mentioned in §3.1 as "topic and focus phrases, sometimes analyzed as separate from clauses"</li> <li>Listed in Table 5 but not exemplified</li> </ul>	<ul style="list-style-type: none"> <li>"John", I saw him" (left-dislocated topic)</li> <li>"I saw him, John" (right-dislocated)</li> <li>"As for Mary, she left early"</li> <li>"Beans, I don't like" (fronted focus)</li> <li>May have special marking (morphemes/intonation)</li> </ul>
Int	Interactive	Heine's (2023) category including interjections, discourse markers, response signals, social formulae, vocatives	Manages discourse flow and interaction rather than propositional content	<ul style="list-style-type: none"> <li>Categories from Heine listed in (11): attention signals (<b>look, hey</b>), directives (<b>hang on</b>), discourse markers (<b>anyway, I mean</b>), evaluatives (<b>good gracious</b>), etc.</li> </ul>	<ul style="list-style-type: none"> <li>"well" in "Well, I think so"</li> <li>"you know" in "I was, <b>you know</b>, tired"</li> <li>"ouch" in "Ouch! That hurt"</li> <li>"hello" in "Hello, how are you?"</li> <li>"Dad" in "Dad, can I go?"</li> <li>May be words or phrases (§2.1, §3.1)</li> </ul>

---

## KEY DESIGN PRINCIPLES UNDERLYING THE LABELS

### 1. Deliberate Under-specification (§3.1)

Croft intentionally does NOT distinguish subtypes:

- **Mod** does NOT differentiate: Article, Demonstrative, Numeral, Adjective
- **Arg** does NOT differentiate: Subject, Object, Oblique (or Ergative, Absolutive, etc.)
- **CPP** does NOT differentiate: Auxiliary, manner adverb, tense marker, etc.

**Rationale:** Simplicity for cross-linguistic application; detailed info recoverable from:

- Glosses (morphological categories in IMT)
- Lexicon (word classes)
- Word order patterns

### 2. Strategy vs. Role Distinction (Table 5)

- **Roles:** Semantic/functional relations (Head, Mod, Pred, Arg, etc.)
- **Strategies:** Formal means of expressing relations (Adp, Lnk, Clf, Idx, Conj)

Strategies may appear at multiple levels but serve the construction at the higher level:

- **Adp** is phrasal CE but marks clausal relations (§3.2, Table 7, footnote \*)
- **Conj** appears at both phrasal and clausal levels

### 3. Three-Tier Labeling (§3.1)

Labels organized by construction size:

- **Phrasal CEs** = "lexical" tier (word-level elements)
- **Clausal CEs** = "phrasal" tier (phrase-level elements + some words)
- **Sentential CEs** = "clausal" tier (clause-level elements + some phrases/words)

This creates a **fixed-depth annotation** compatible with tabular formats.

## 4. Pragmatic vs. Morphosyntactic Basis (§4.1)

Key theoretical point:

- **Main vs. subordinate** = pragmatic distinction (asserted vs. non-asserted), NOT purely syntactic
- Same morphosyntax may serve different pragmatic functions (insubordination, ex. 24-26)
- Same pragmatic function may use different morphosyntax (coordinate vs. subordinate, ex. 22)

Labels reflect **functional categories** more than formal ones.

## 5. Cross-linguistic Applicability

Labels designed to apply across languages:

- Avoid English-specific categories (e.g., no "article" subtype, since not all languages have articles)
  - Use comparative concepts (Haspelmath 2010) that can be identified cross-linguistically
  - Rely on universal or near-universal distinctions (predication, reference, modification)
- 

# ANNOTATION WORKFLOW WITH LABELS

## Input Text:

The tree that fell on our house also wrecked your car.

## Step 1: Boundary Markers (§2)

The tree + {that + fell + on {our} house} also + wrecked + your + car .

## Step 2: Three-Tier CE Labels (§3)

Analyzed Text: The tree + {that + fell + on {our} house} also + wrecked + your + car

.

Phrasal CEs: Mod Head Head Head Adp Head Head Head Head Head Head Head

d

Clausal CEs: Arg Conj Pred Arg Gen CPP Pred Gen Arg

Sentential CEs: Main

Rel

Main

This table format (compatible with CLDF) provides:

- **Row 1:** Word-level segmentation with boundary markers
  - **Row 2:** Phrasal-level roles (what role does each word play in its phrase?)
  - **Row 3:** Clausal-level roles (what role does each phrase play in its clause?)
  - **Row 4:** Sentential-level roles (what role does each clause play in the sentence?)
- 

## AUTOMATIC ANNOTATION CUES (§3.2, Tables 6-8)

### For Phrasal CEs:

Label	Detection Cues
Head	Has number/gender/definiteness/case morphology; content word class
Mod	Closed-class (det, dem, num); adjective class; word order (Mod typically before Head in English)
Adm	Degree word class; occurs before Mod in English (Adm-Mod-Head order)
Adp	Adposition word class; precedes/follows NP
Lnk	Linking particle/affix; language-specific
Clf	Classifier word class; follows/precedes numeral
Idx	Index word (rare); agreement morphology
Conj	Conjunction word class; occurs between coordinated elements

### For Clausal CEs:

Label	Detection Cues
Pred	TAM morphology; verb word class; required element
Arg	Case/adposition marking; word order (SOV, SVO patterns); noun phrase
CPP	Auxiliary word class; modal class; aspect marker; adverb class
Gen	Genitive case/adposition; possessive construction; word order (GenN vs NGen)
FPM	Spatial/temporal case/adposition; locative/temporal marker
Conj	Coordinating conjunction; occurs between clauses

### For Sentential CEs:

Label	Detection Cues

<b>Main</b>	No subordination marking; indicative mood; root clause inflection
<b>Adv</b>	Adverbializer; converb morphology; subordinating conjunction (when, because, if)
<b>Rel</b>	Relativizer (that, who); participle morphology; relative clause marker
<b>Comp</b>	Complementizer (that, whether); infinitive; action nominal; complement-taking predicate in matrix
<b>Dtch</b>	Topic/focus marker; special intonation/word order; pragmatically detached
<b>Int</b>	Interjection class; discourse marker class; ideophone; vocative case

---

This comprehensive table should provide the depth needed to understand and potentially implement Croft's labeling system computationally.