

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
Head	Head	The core 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"> "brothers" in "The two brothers" "lose" in "might lose" "competition" in "the competition" 	<ul style="list-style-type: none"> "dog" in "the big dog" "quickly" in "very quickly" Single-word phrases are annotated as Head (§3.1)
Mod	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"> "The two" in "The two brothers" "the" in "the game" "a rare but aggressive" in "a rare but aggressive cancer" (Table 3) 	<ul style="list-style-type: none"> "big" in "the big dog" "this" in "this book" "three" in "three cats" Subtypes not distinguished: Art, Dem, Num, Adj all labeled Mod (§3.1)
Adm	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"> "very" in "a very large tree" (Table 2) "almost" in "almost fifty" (§2.1) 	<ul style="list-style-type: none"> "extremely" in "extremely hot" "quite" in "quite interesting" "somewhat" in "somewhat difficult" No separate attributive phrase level created (§2.1)
Adp	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"> "on" in "on top of" (before becoming MWE) "on" in "fell on {our} house" (Table 4) 	<ul style="list-style-type: none"> English: in, at, from, to Turkish: ev-de 'house-at' (postposition) Found in phrases but indicate clause-level CEs (§3.2, Table 7)
Lnk	Linker	Grammatical elements connecting modifiers to heads; linking particles or morphemes	Strategy for marking modification relationships	<ul style="list-style-type: none"> Spanish "de" in possessive constructions Tagalog linker particles (implied from typological context) 	<ul style="list-style-type: none"> Mandarin "de" (的): hóng de shū 'red DE book' Malay "yang": rumah yang besar 'house THAT big' Found in both phrasal and clausal contexts (Tables 6-7)
Clf	Classifier	Numeral or noun classifiers categorizing referents by shape, animacy, or other properties	Strategy for expressing quantification or categorization	<ul style="list-style-type: none"> Not exemplified in English examples Referenced in typological discussion (§3.1) 	<ul style="list-style-type: none"> Mandarin: sān bēn shū 'three CLF book' Thai: rôt sām khan 'car three CLF' Japanese: hon ni-satsu 'book two-CLF'
Idx	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"> Rare in English Implied in discussion of "indexation" throughout §3.2 	<ul style="list-style-type: none"> Swahili: wa-toto wa-kubwa 'children big' (both have CL2 prefix) Usually bound morphemes, but labeled when orthographically separate (§3.1)

Conj	Conjunction (Phrasal)	Coordinators linking modifiers or phrases within a larger phrase structure	Strategy for coordinating elements at phrase level	<ul style="list-style-type: none"> • "and" in "the adults and the kids" (Table 3) • "but" in "a rare but aggressive cancer" (Table 3) 	<ul style="list-style-type: none"> • "or" in "tea or coffee" • "both...and" in "both the priest and the shopkeeper" (ex. 2) • Comma in lists: "red, white, and 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
Pred	Predicate	The primary predicate of the clause; the main verb or predicate element expressing the event/state	Core semantic element denoting the situation	<ul style="list-style-type: none"> • "lose" in "might + lose" (Table 1) • "wrecked" in "also + wrecked + your + car" (Table 4) • "fell" in "{that + fell + on {our} house}" (Table 4) 	<ul style="list-style-type: none"> • "runs" in "The dog runs" • "is" in "Ivan is the best dancer" (ex. 5b) • In complex predicates, the primary predicate only (§3.1)
Arg	Argument	Referring phrases denoting participants in the event; includes subjects, objects, and obliques—NOT differentiated into subtypes	Expresses participants/entities involved in the situation	<ul style="list-style-type: none"> • "The two brothers" in "The two brothers + might + lose" (Table 1) • "the game" in "lose + the game" (Table 1) • "the competition" in "exit + the competition" (Table 1) 	<ul style="list-style-type: none"> • "She" in "She ate" • "the book" in "read the book" • "to school" in "walked to school" • Broad definition includes "adjuncts" (§3.1)
CPP	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	<ul style="list-style-type: none"> • "might" in "might + lose" (Table 1) • "also" in "also + wrecked" (Table 4) • "had" in "had + died" (implied from text) 	<ul style="list-style-type: none"> • "can" in "can run" • "will" in "will go" • "been" in "has been eating" • "carefully" in "drives carefully" (adverb) • Not grouped into predicate phrases due to non-contiguity (§3.1)
Gen	Genitive Phrase	Possessive or genitive phrases modifying another nominal referent	Expresses possessive relationships between referents	<ul style="list-style-type: none"> • "our" in "on {our} house" (Table 4) • "your" in "wrecked + your + car" (Table 4) • "the dog's" in "the dog's + toy" (ex. 15a) 	<ul style="list-style-type: none"> • "Mary's" in "Mary's book" • "of the table" in "the surface of the table" (§4.2.2) • May be single words (pronouns) or full phrases (§3.2)
FPM	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	<ul style="list-style-type: none"> • Implied in discussion of "adpositional and case-marked noun modifiers" (§3.1) • Contrasted with Gen in Table 5 	<ul style="list-style-type: none"> • "on the table" in "the book on the table" • "from Boston" in "the train from Boston" • Turkish: masa-da-ki kitap 'table-LOC-REL book' ("book on the table") • Grouped with Gen as "flags" in typology (§3.2, Table 7)

Conj	Conjunction (Clausal)	Coordinators or linking elements joining clauses	Strategy for connecting clauses	<ul style="list-style-type: none"> • "and" in "lose + the game # and + exit" (Table 1, ex. 5c) • "and" in "buy and eat" (ex. 5c) 	<ul style="list-style-type: none"> • "but" in "I tried but I failed" • "or" in "Stay or go" • "so" in "exhausted and (so) 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
Main	Main Clause	Clauses expressing pragmatically asserted events/states; independent clauses that could stand alone	Primary assertion in the utterance	<ul style="list-style-type: none"> • Both clauses in coordinate structure: "lose the game # and + exit the competition" (both labeled Main in Table 1) 	<ul style="list-style-type: none"> • "I ate" in "I ate and left" • "She left" in "She left" • May occur with subordinate clauses: "I left when he arrived" • Distinguished by assertion, not morphosyntax alone (§4.1)
Adv	Adverbial Clause	Subordinate clauses expressing temporal, causal, conditional, or other circumstantial relations to main event	Provides circumstantial/temporal/causal background	<ul style="list-style-type: none"> • "after + I + ate" in "after + I + ate # I + left" (ex. 14a) • "If you do that" in "If you do that, the terrorists have won" (ex. 22d) 	<ul style="list-style-type: none"> • "when he arrived" in "I left when he arrived" • "because she was exhausted" in "She went to bed because she was exhausted" (ex. 22b) • "Although John had no money" in "Although John had no money, he went into this restaurant" (ex. 22e)
Rel	Relative Clause	Clauses functioning as modifiers of nominal referents; attributive clauses	Restricts or describes the referent of a noun phrase	<ul style="list-style-type: none"> • "that + fell + on {our} house" in "The tree {that + fell + on {our} house} had + died" (Table 4) • "that + Gollum + had + lost" in "the ring # that + Gollum + had + lost" (ex. 16) 	<ul style="list-style-type: none"> • "who is tall" in "the man who is tall" • "that I bought" in "the bicycle that I bought" (§3.2) • May interrupt matrix clause (center-embedding) or follow head (§4.2.1)
Comp	Complement Clause	Clauses functioning as arguments of predicates; complete propositions filling argument slots	Serves as subject, object, or complement of matrix predicate	<ul style="list-style-type: none"> • "that + he + resigned" in "It + isn't + surprising # that + he + resigned" (ex. 35b) • "to believe the girl" in "wants to believe the girl" (ex. 6, AMR example) 	<ul style="list-style-type: none"> • "that I bought a bicycle" in "I told her that I bought a bicycle" (§3.2) • "to leave" in "She wants to leave" • 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"> • Mentioned in §3.1 as "topic and focus phrases, sometimes analyzed as separate from clauses" • Listed in Table 5 but not exemplified 	<ul style="list-style-type: none"> • "John", I saw him" (left-dislocated topic) • "I saw him, John" (right-dislocated) • "As for Mary, she left early" • "Beans, I don't like" (fronted focus) • May have special marking (morphemes/intonation)
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"> • Categories from Heine listed in (11): attention signals (look, hey), directives (hang on), discourse markers (anyway, I mean), evaluatives (good, gracious), etc. 	<ul style="list-style-type: none"> • "well" in "Well, I think so" • "you know" in "I was, you know, tired" • "ouch" in "Ouch! That hurt" • "hello" in "Hello, how are you?" • "Dad" in "Dad, can I go?" • May be words or phrases (§2.1, §3.1)

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

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
-------	----------------

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