

Why the Copula Appears: The Case of Predicate Type and Copula Selection in Thai

Piyapath T Spencer
Universität Konstanz

Thai nominal copular clauses show a three-way alternation: zero predication ($\emptyset$), characterisational *pen*, and identificational $k^h\text{ur}$. The semantic contrast between the overt copulas is documented in the Thai descriptive literature (Kuno & Wongkhamthong, 1981; Komolwanig & Sawada, 1993) but has not been connected to the broader typology of copular clause subtypes (Higgins, 1979; Mikkelsen, 2005; den Dikken, 2006). The distribution of $\emptyset$ has been treated as underlying verbal structure (Warotamasikkhadit, 1969) or, within Lexicase, as nominal predicates heading the clause directly without null verbal elements (Indrambarya, 2000); neither account extends to a formal treatment of the *pen*– $k^h\text{ur}$ distinction. I argue that $\emptyset$ is the unmarked default, licensed when the predicate nominal denotes an individual (Partee, 1987) rather than a property, and that the predicational–specificational distinction (Mikkelsen, 2005; Percus & Sharvit, 2024) is encoded in the lexical semantics of the copular items themselves, which is irreducible to information structure alone. That is, *pen* heads predicational clauses, whereas $k^h\text{ur}$ heads specificational/equative clauses.

Data. Temporal nominals permit all three forms; role nominals prefer *pen* and reject $\emptyset$:

- (1) a. wan ní: { $\emptyset$ / *pen* / $k^h\text{ur}$ } wan-sǒŋ.kra:n
day this cop day-Songkran
'Today is Songkran day.' $\emptyset > \textit{pen} > k^h\text{ur}$
- b. $k^h\text{aw}$ *(*pen* / ?? $k^h\text{ur}$) nak-rian
he cop student
'He is a student.' * $\emptyset$, *pen*, ? $k^h\text{ur}$

Temporal nominals denote individual concepts requiring no assertive categorisation; role nominals denote properties ($\langle e, t \rangle$) and require an overt copular element (Dalrymple et al., 2004; Nordlinger & Sadler, 2006). Five tests separate *pen* from $k^h\text{ur}$, each deriving from the property-predication/identity-equation contrast (Schlenker, 2003; Percus & Sharvit, 2024):

Test	$\emptyset$	<i>pen</i>	$k^h\text{ur}$	Explanation
Negation / modals	✓	✓	*	Identity admits no epistemic hedging
Duration adverbial	—	✓	*	Identity is not temporally bounded
Indefinite complement	✓	✓	*	Equation requires a unique referent
Subject-compl. inversion	—	*	✓	Equation is symmetric; predication is not
Answer to <i>which</i> ?	✓	*	✓	$k^h\text{ur}$ saturates an open focus variable

Subject-complement inversion is grammatical with $k^h\text{ur}$ but not with *pen* as expected when $k^h\text{ur}$ heads specificational clauses whose subjects denote properties rather than individuals (Mikkelsen, 2005):

- (2) lû:k-sǎ:w khun-sǒm.sĩ: $k^h\text{ur}$ /**pen* dèk-khon-nán
daughter Mrs.Somsri cop child-CL-that
'Mrs Somsri's daughter is that child.'

A cross-clausal construction further differentiates in that $k^h\text{ur}$ resists subsequent independent predication on the same discourse referent, while *pen* and $\emptyset$ do not:

- (3) dèk-khon-nán *pen* / $\emptyset$ / # $k^h\text{ur}$ lû:k-sǎ:w khun-sǒm.sĩ:, kɛ̀ khà.yǎn di:
child-CL-that cop daughter Mrs.Somsri she diligent good
'That child is Mrs Somsri's daughter; she is very diligent.'

Such discourse closure follows from the identity semantics of $k^h\text{ur}$: once a referent is equated with another individual, subsequent independent predication becomes referentially redundant.

Information structure. The predication/specificational distinction has been analysed as an information-structural distinction in which predication clauses have a TOPIC subject and FOCUS predicate, marking the constituent contributing the at-issue alternative (Rooth, 1992), while specificational clauses have a focus-variable subject (Mikkelsen, 2005; Heycock & Kroch, 2002). This characterisation is adopted here, but information structure is treated as derivative of the underlying semantic distinction rather than its source: modal blocking, duration incompatibility, and the discourse closure in (3) all follow from the identity semantics of k^h_{ui} : without additional information-structural machinery, and the information-structural configurations of *pen* and k^h_{ui} are therefore encoded as constraints on PRED in f-structure rather than independently specified.

Analysis. Within LFG (Bresnan et al., 2016; Dalrymple, 2001), $\emptyset$ constructions receive a *single-tier* analysis (Dalrymple et al., 2004; Nordlinger & Sadler, 2006) which allows the nominal to contribute the clausal PRED directly, and its individual-concept denotation $\langle s, e \rangle$ makes it a self-standing predicate rather than an argument of a further relation. Overt copula constructions receive a *double-tier* analysis in which the copula contributes a relational PRED over SUBJ and PREDLINK; k^h_{ui} 's PREDLINK, though also individual-denoting (e), is an *argument* of the relation EQ, not a clause-heading predicate, distinguishing it structurally from $\emptyset$ despite the shared semantic type. The distinction among $\emptyset$, *pen* and k^h_{ui} is encoded directly in the PRED value: in $\emptyset$, the nominal's own PRED subcategorizes for SUBJ with no further embedding; in *pen* and k^h_{ui} , by contrast, the nominal is not the clausal PRED at all but a PREDLINK argument nested inside a relation the copula introduces.

<i>wan-sǒŋ.krɪn</i> 'Songkran day' ($\emptyset$)	k^h_{aw} <i>pen nak-rian</i> 'He is a student'	$k^h_{aw} k^h_{ui}$ <i>nak-rian khon-nán</i> 'He is that student'
$\left[\begin{array}{c} \text{PRED 'SONGKRAN-DAY (SUBJ)'} \\ \text{SUBJ} \left[\begin{array}{c} \text{PRED 'DAY'} \\ \text{DEF +} \\ \text{SEM-TYPE } \langle s, e \rangle \end{array} \right] \end{array} \right]$	$\left[\begin{array}{c} \text{PRED 'PEN (SUBJ, PREDLINK)'} \\ \text{SUBJ} \left[\begin{array}{c} \text{PRED 'HE'} \\ \text{TOPIC +} \\ \text{PRED 'STUDENT'} \\ \text{DEF -} \\ \text{FOCUS +} \\ \text{SEM-TYPE } \langle e, t \rangle \end{array} \right] \\ \text{PREDLINK} \left[\begin{array}{c} \text{DEF -} \\ \text{FOCUS +} \\ \text{SEM-TYPE } \langle e, t \rangle \end{array} \right] \end{array} \right]$	$\left[\begin{array}{c} \text{PRED 'EQ (SUBJ, PREDLINK)'} \\ \text{SUBJ} \left[\begin{array}{c} \text{PRED 'HE'} \\ \text{FOCUS + [var]} \\ \text{PRED 'THAT.STUDENT'} \\ \text{DEF +} \\ \text{SEM-TYPE } e \end{array} \right] \\ \text{PREDLINK} \left[\begin{array}{c} \text{DEF -} \\ \text{FOCUS +} \\ \text{SEM-TYPE } e \end{array} \right] \end{array} \right]$
Individual concept heads the clause directly as PRED; no PREDLINK, no relational PRED.	Property pred.; IS: TOP subj, FOC predlink. Modals and duration permitted.	Identity/equation; IS: FOC-var subj. Modals and further predication blocked.

The PRED value EQ encodes the asymmetric identity relation of Schlenker (2003) and Percus & Sharvit (2024): the PREDLINK denotes an individual (e) while the SUBJ functions as an individual concept, accounting for the definiteness requirement, the licensing of inversion through semantic symmetry, and the discourse closure effect. The property-denoting PREDLINK of PEN licenses temporal modification and modals. All distributional contrasts follow without construction-specific stipulation.

Implications. Zero predication in nominal copular clauses is cross-linguistically widespread (Stassen, 1997, 2008; Pustet, 2003), and the literature has proposed null copular elements to account for it across frameworks: a null T head governing tense in Arabic (Benmamoun, 2000), a null predicate head in Romance (Moro, 1997), and a null PRED contributed by phrase structure annotation in copula-less languages (Nordlinger & Sadler, 2006; Attia, 2008). These analyses share the assumption that $\emptyset$ is the form requiring explanation. The Thai data, however, suggest once the semantic conditioning of $\emptyset$ by predicate nominal type is identified, the postulation of null elements is unmotivated. It is the overt copulas that require explanation, and that explanation which is grounded in the predication/specificational distinction demonstrates that the distinction is encoded in the lexical semantics of copular items rather than solely through information structure or prosody. It is thus a fundamental mechanism of clausal semantics in languages without case, gender, or morphological inflection.

Selected References: Attia, M. 2008. A unified analysis of copula constructions in LFG. In *Proceedings of LFG08*. • Bresnan, J., Asudeh, A., Toivonen, I., & Wechsler, S. 2016. *Lexical-Functional Syntax*. • Dalrymple, M., Dyvik, H., & King, T. H. 2004. Copular complements: Closed or open? In *Proceedings of LFG04*. • den Dikken, M. 2006. *Relators and Linkers: The Syntax of Predication, Predicate Inversion, and Copulas*. • Heycock, C., & Kroch, A. 2002. Inversion and equation in copular sentences. In *ZAS Papers in Linguistics* 29. • Higgins, F. R. 1979. *The Pseudo-Cleft Construction in English*. • Indrambarya, K. 2000. On nonverbal predicates in Thai. • Komolwanig, K., & Sawada, N. 1993. A contrastive study of copulative sentences in Japanese and Thai. • Kuno, S., & Wongkhamthong, P. 1981. Two copulative verbs in Thai. • Mikkelsen, L. 2005. *Copular Clauses: Specification, Predication and Equation*. • Moro, A. 1997. *The Raising of Predicates: Predicative Noun Phrases and the Theory of Clause Structure*. • Partee, B. 1987. Noun phrase interpretation and type-shifting principles. • Pustet, R. 2003. *Copulas: Universals in the Categorization of the Lexicon*. • Schlenker, P. 2003. Clausal equations. *NLLT* 21(1). • Stassen, L. 1997. *Intransitive Predication*. • Warotamasikkhadit, U. 1969. *Thai Syntax: A Generative Study*.