

Cross-linguistic Analysis of the Mermaid Constructions in LFG

Chihiro Taguchi

University of Notre Dame
<https://ctaguchi.github.io>

October 29, 2022
33rd South of England LFG Meeting

Outline

Introduction

Descriptive Analysis

LFG Analysis

Cross-linguistic Analysis

Conclusion

In this talk...

Objective

- To analyze the Mermaid Constructions across languages with LFG

Background (from LFG 2022 at Groningen)

- “Mermaid Constructions in LFG”
- Takeaways: Mermaid Constructions are **control** and **raising** with a **nominal predicate**
- Evidence was mainly data from Japanese

What's new in this talk?

- Provide a cross-linguistic account

Examples (contd.)

- (5) Amdo Tibetan (< Tibeto-Burman < Sino-Tibetan; SOV)⁸

arʃa=kə [*nor pʈsoŋ-ʃu*] *ntɕʰarʒə* *rɛ*
 father=ERG yak sell.IPFV-NMLZ.GEN plan COP.B

‘(My) father plans to sell yaks’

- (6) nDrapa (Zhaba) (< Qiangic < Sino-Tibetan; SOV)⁹

ɲoro1 [*kaotɔŋ1 lo=ɕi2 fidi=tʃi3*] *nkhei1* *rɛ*
 3SG high_school learn=wish think=IPFV appearance COP

‘He seems to want to go to high school’

- (7) Burmese (< Tibeto-Burman < Sino-Tibetan; SOV)¹⁰

tù [*dì híN=gò sá=dê*] *póun(=bé)*
 3SG this curry=KO eat=ADN shape(=EMP)

‘(S)he seems to have eaten this curry’

B — pattern B (see Ebihara 2020), EMP — emphasis, KO — *kò/gò* (see Kato 2020)

⁸Shiho Ebihara. “Amdo Tibetan”. In: *Mermaid Construction: A Compound-Predicate Construction With Biclausal Appearance*. Ed. by Tasaku Tsunoda. Comparative Handbooks of Linguistics. Mouton de Gruyter, 2020, pp. 419–464.

⁹Satoko Shirai. “nDrapa”. In: *Mermaid Construction: A Compound-Predicate Construction With Biclausal Appearance*. Ed. by Tasaku Tsunoda. Comparative Handbooks of Linguistics. Mouton de Gruyter, 2020, pp. 465–509.

¹⁰Atsuhiko Kato. “Burmese”. In: *Mermaid Construction: A Compound-Predicate Construction With Biclausal Appearance*. Ed. by Tasaku Tsunoda. Comparative Handbooks of Linguistics. Mouton de Gruyter, 2020, pp. 781–816.

Examples (contd.)

- (8) Kurukh (< Northern Dravidian < Dravidian; SOV)¹¹

a:s-hi: [tamba-s-in ilc-ka:] *caɖɖe:* rahc-a:

3SG.M-GEN own.father-M-ACC fear.PS-PST.VADJ necessity COP.PS-PST.3SG.NM

‘It was because he was afraid of his father’

- (9) Sidaama (< Cushitic < Afro-Asiatic; SOV)¹²

íse [faraššó guluf-f-annó] *gara-a=ti*

3SG.F.NOM horse.ACCOBL ride-3SG.F-IPFV.3 manner-LV=NPC.PRED.MOD

‘She seems to ride a horse (habitually)’

- (10) Bengali (< Indo-Iranian < Indo-European; SOV)¹³

tar [tokio=te jawar] *kɔt^ha*

3SG.GEN Tokyo=LOC go.NMLZ.GEN word

'It is planned that he s/he is going to Tokyo'

ACCOBL — Accusative-Oblique, LV — lengthened vowel, MOD — modified, NM — non-masculine, NPC — nominalized predicate clitic, PRED — predicative, PS — past stem, VADJ — verbal adjective

¹¹ Masato Kobayashi and Tasaku Tsunoda. “Kurux”. In: *Mermaid Construction: A Compound-Predicate Construction With Biclausal Appearance*. Ed. by Tasaku Tsunoda. Comparative Handbooks of Linguistics. Mouton de Gruyter, 2020. pp. 781–816.

¹² Kazuhiro Kawachi. "Sidaama". In: *Mermaid Construction: A Compound-Predicate Construction With Biclausal Appearance*. Ed. by Tasaku Tsunoda. Comparative Handbooks of Linguistics. Mouton de Gruyter, 2020. pp. 781–816.

¹³Keisuke Huziwaru. "A contrastive study of external adnominal clauses in Japanese and Bangla". In: *Bangabidya: International Journal of Bengal Studies* 10 (2018), pp. 358–367.

Examples: Different case-marking in Tatar

(14) Tatar (< Kipchak < Turkic; SOV)¹⁵

siŋa [joqla-rya] *röxsät*
2SG.DAT sleep-INF permission

‘You are allowed to sleep.’

(15) Tatar (< Kipchak < Turkic; SOV)¹⁶

Marat-niŋ [joqla-rya] *isäb-e*
Marat-GEN sleep-INF thought-POSS.3

‘Marat is going to sleep.’

Dative vs. Genitive (+ possessive agreement)

¹⁵ Chihiro Taguchi. “Mermaid Construction in Tatar”. In: *Proceedings of the 162nd Conference of the Linguistic Society of Japan*. 2021.

¹⁶ Chihiro Taguchi. “Mermaid Construction in Tatar”. In: *Proceedings of the 162nd Conference of the Linguistic Society of Japan*. 2021. ▶

¹⁷ Shaimaa ElSadek and Louisa Sadler. In: *Proceedings of the LFG15 Conference*. CSLI Publications, 2015. »

Examples against the compound predicate analysis

Tsunoda (2020) says adjacent V and N become a compound predicate, but...

(18) Russian (< Slavic < Indo-European; SVO)

pora *nam* [*uxodit'*]
time 1PL.DAT PF:leave:INF

‘It is time for us to leave’

(19) Welsh (< Celtic < Indo-European; VSO)

Rhaid *i* *fi* [*godi'n* *gynnar*]
rhaid *i* *fi* [*godi* *yn* *gynnar*]
necessity to 1SG wake_up.INF in early

‘I need to wake up early’

- N_{MC} and V are not adjacent to each other in these languages
- How can they form a compound predicate? (they don't)

Typology of Mermaid Constructions

Case-marking of subject/agent

Case	Languages
NOM/ERG	Japanese, Mitsukaido, Irabu, Korean, Amdo-Tibetan, nDrapa, Burmese, Sidaama, Tagalog
GEN/POSS	Kurukh, Bengali, Tagalog, Tatar, Egyptian Arabic
DAT/to	Tatar, Russian, Welsh

Form of the embedded predicate

Function	Form	Languages
Adnominal	ADN NMLZ.GEN VADJ	Japanese, Mitsukaido, Irabu, Korean, nDrapa, Burmese, Sidaama Amdo-Tibetan, Bengali Kurukh
Infinitive		Tagalog, Tatar, Russian, Welsh
Finite		Tagalog, Egyptian Arabic

... Quite diverse

Tsunoda's (2020)¹⁸ generalization:

- MCs are “in the main confined to Asia” ... not really

¹⁸Tasaku Tsunoda. *Mermaid Construction: A Compound-Predicate Construction With Biclausal Appearance*. *Comparative Handbooks of Linguistics*. Mouton de Gruyter, 2020.

Proposal

Proposal (from LFG22)

- MC is **anaphoric/functional control** with a nominal predicate
- MC does **not** involve a compound predicate
- Therefore, MC is **biclausal**
- MC can be analyzed in LFG just like other anaphoric/functional control

Proposal (new)

- Typological diversity of MCs can be captured in f-structure
- different case-marking: subcategorization of predicates
- different verb forms:

Subject is in the matrix clause: Evidence 1

The following slides demonstrate that an MC subject is in the matrix clause.

- A comparison with the similar non-MC (21)
- In (21), *Hanako(=ga)* is inside the modifier clause, so an additional subject argument can be added in the matrix clause (otherwise *pro*).
- (20), in contrast, doesn't allow it
- the matrix subject is already **occupied**

(20) Japanese (MC)

(**kore=ga*) *Hanako=ga sakana=o yak-u yotei=da*
 this=NOM Hanako=NOM fish=OBJ grill-NPST.ADN plan(N)=COP

‘Hanako is going to grill the fish.’

(21) Japanese (non-MC)

(*kore=ga*) [*Hanako=ga sakana=o yak-u*] *nioi=da*
 this=NOM Hanako=NOM fish=ACC grill-PRS.ADN smell(N)=COP

‘(This is) the smell that [comes from where] Hanako grills the fish.’

Syntactic tests: Passivization

- Functional control (29), (30): passivization does not change thematic relation
- Anaphoric control (31), (32): passivization **does** change thematic relation

(29) a. Tom seems to hit Jerry. *seem(hit(Tom, Jerry))*
 b. Jerry seems to be hit by Tom. *seem(hit(Tom, Jerry))*

(30) a. Japanese (functional control)
Hanako=*ga*_i [_{t_i} *Taroo*=*o* *tatak-u*] *yotei*=*da*
 Hanako=NOM Taro=ACC hit-NPST.ADN plan(N)=COP
 ‘Hanako is going to hit Taro: *planned(hit(h, t))*’

b. Japanese (functional control, passivized)
Taroo=*ga*_i [_{t_i} *Hanako*=*ni* *tatak-are-ru*] *yotei*=*da*
 Taro=NOM Hanako=DAT hit-PASS-NPST.ADN plan(N)=COP
 ‘Taro is going to be hit by Hanako: *planned(hit(h, t))*’

- *yotei* shares a characteristic of a functional control predicate

Syntactic tests: Idiom chunks

- Functional control (33 b), (34 b) keeps idiomatic meaning
- Anaphoric control only allows for literal meaning

- (33) a. The cat is out of the bag. (i.e., the secret is revealed)
b. The cat seems to be out of the bag.

- (34) a. Japanese

asi=ga boo=ni nar-u

leg=NOM stick=DAT become-NPST

‘The legs become sticks (i.e., exhausted)’

- b. Japanese

asi=ga boo=ni nar-u yotei=da

leg=NOM stick=DAT become-NPST.ADN plan(N)=COP

‘The legs are going to be sticks (i.e., exhausted)’

- *yotei* shares a characteristic of a functional control predicate

Syntactic tests: Idiom chunks

- Functional control keeps idiomatic meaning
- Anaphoric control (35 b), (36 b) only allows for literal meaning

(35) a. The cat is out of the bag.

b. # The cat tries to be out of the bag.

(36) a. Japanese

asi=ga boo=ni nar-u

leg=NOM stick=DAT become-NPST

‘The legs become sticks (i.e., exhausted)’

b. Japanese

asi=ga boo=ni nar-u ki=da

leg=NOM stick=DAT become-NPST.ADN plan(N)=COP

‘The legs intend to become sticks’

- *ki* shares a characteristic of an anaphoric control predicate

(Hopefully theory-neutral) Interim summary

Tsunoda (2020)'s Analysis of MC

- V and N_{MC} compose a compound predicate
- MC is monoclausal

We have seen:

- MC can be treated as **anaphoric/functional control**
 - but with a **noun predicate**
- It follows that MC is **biclausal**

Next:

- Analyze MC in LFG

Motivation for an LFG analysis

- PRED readily allows for a nominal predicate
- Therefore, lexical entries of MC nouns (N_{MC}) have a similar form to anaphoric/functional control
 - Anaphoric control MC:
 - $(\uparrow \text{PRED}) = 'N_{MC} \langle (\text{SUBJ} | \text{OBL}_{\theta}), \text{COMP} \rangle^{21}$
 - $(\uparrow \text{COMP SUBJ PRED}) = \text{'PRO'}$
 - Functional control MC:
 - $(\uparrow \text{PRED}) = 'N_{MC} \langle \text{XCOMP} \rangle \text{SUBJ}$
 - $(\uparrow \text{SUBJ}) = (\uparrow \text{XCOMP SUBJ})$
- COMP/XCOMP can readily handle the cross-linguistic variation of non-finite forms
 - Infinitive: Tagalog, Tatar, Russian, Welsh
 - Adnominal/verbal adjective: Japanese, Burmese, Kurukh, Sidaama, etc.
 - Verbal noun: Amdo Tibetan, Bengali
- Severing syntax, function, and semantics (c/f/s-structure)
 - Disentangling MC's mystery: 'Syntactically nominal, functionally predicative, and semantically abstract (modal, etc.)?'

²¹ Whether SUBJ or OBL_θ is selected depends on each lexeme.

c-structure of functional control with N_{MC}

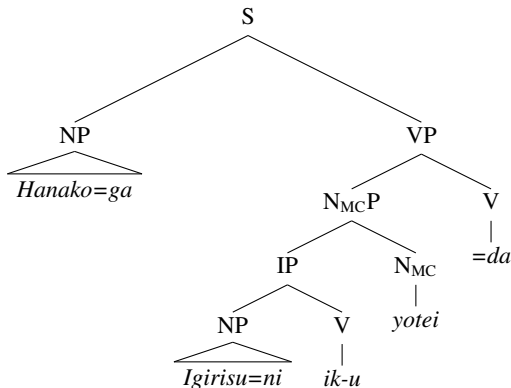
(37) Japanese (< Japonic)

Hanako=ga Igirisu=ni ik-u yotei=da

Hanako=NOM UK=DAT go-NPST.ADN plan(N)=COP

‘Hanako is going to go to the UK.’

(38)



c-structure of anaphoric control with N_{MC}

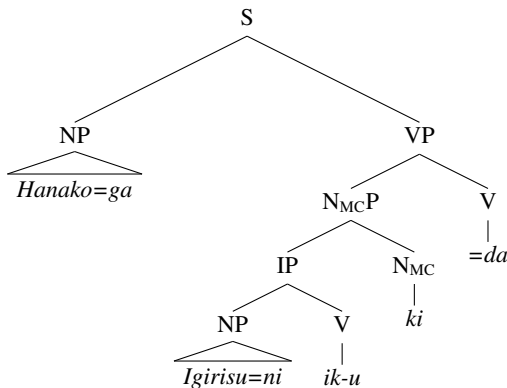
(39) Japanese (< Japonic)

Hanako=ga Igirisu=ni ik-u ki=da

Hanako=NOM UK=DAT go-NPST.ADN feeling(N)=COP

‘Hanako intends to go to the UK.’

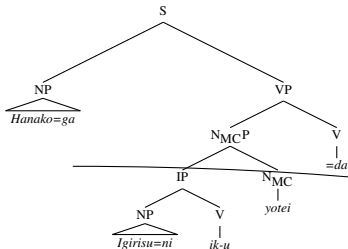
(40)



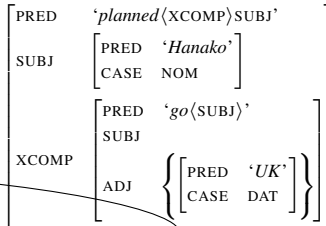
- c-structural form is the same in raising and equi (language-specific)

f-structure of functional control with N_{MC}

(41)

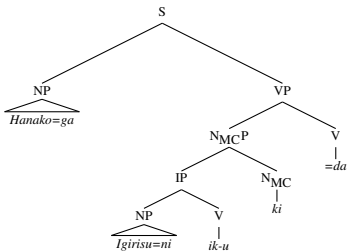


(42)

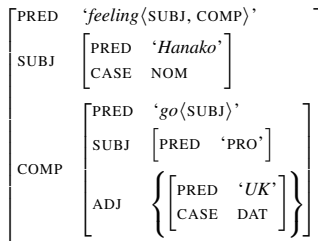


f-structure of anaphoric control with N_{MC}

(43)



(44)



- Raising and equi MCs have a different f-structure (lexical entries)

Semantics of the Mermaid Constructions

- The meaning of MCs can also be constructed in the same manner as functional/anaphoric control
- Following the semantic representation of raising and equi verbs by Dalrymple et al. (2019)²²:
 - David seemed to yawn: *seem*(yawn(*David*))
 - David tried to yawn: *try*(*David*, yawn(*David*))
- We expect the same representation for MCs too:
 - Hanako=ga Nagoya=ni ik-u yotei=da: *planned*(go(*Hanako*, UK))
 - Hanako=ga Nagoya=ni ik-u ki=da: *feeling*(*Hanako*, go(*Hanako*, UK))
- Meaning constructors for N_{MC} :
 - raising: $\lambda P.yotei(P) : (\uparrow \text{XCOMP})_\sigma \multimap \uparrow_\sigma$
 - equi: $\lambda P \lambda x.ki(x, P(x)) :$
 $((\uparrow \text{COMP SUBJ})_\sigma \multimap (\uparrow \text{COMP PRED})_\sigma) \multimap ((\uparrow \text{SUBJ})_\sigma \multimap \uparrow_\sigma)$

²²Mary Dalrymple, John J. Lowe, and Louise Mycock. *The Oxford Reference Guide to Lexical Functional Grammar*. Oxford University Press, 2019.

Predicting the scope of anaphoric control

- Anaphoric control analysis correctly restricts the scope

(50) Japanese
dareka=ga ik-u ki=da
 someone=NOM go-NPST.ADN feeling(N)=COP
 ‘Someone intends to go. (wide: $\exists > \text{intend}$; *narrow: $\text{intend} > \exists$)’

(51)
$$i \left[\begin{array}{l} \text{PRED} \quad \text{'intend' (SUBJ, COMP)} \\ \text{SUBJ} \quad \left[\begin{array}{l} \text{PRED} \quad \text{'someone'} \\ \text{CASE} \quad \text{NOM} \end{array} \right] \\ \text{COMP} \quad \left[\begin{array}{l} \text{PRED} \quad \text{'go' (SUBJ)} \\ \text{SUBJ} \quad \text{'PRO'} \end{array} \right] \end{array} \right]_g$$

(52)
$$\begin{array}{lll} [\text{ki}] & \lambda P \lambda x. \text{intend}(x, P) & : (s_\sigma \multimap g_\sigma) \multimap (s_\sigma \multimap i_\sigma) \\ [\text{dareka}] & \lambda S. \text{exist}(x, \text{person}(x), S(x)) & : \forall H. (s_\sigma \multimap H) \multimap H \\ [\text{iku}] & \lambda x. \text{go}(x) & : s_\sigma \multimap g_\sigma \end{array}$$

(53) Wide:
$$\begin{array}{lll} [\text{iku-ki}] & \lambda x. \text{intend}(\text{go}(x)) & : s_\sigma \multimap i_\sigma^{24} \\ [\text{dareka-iku-ki}] & \text{exist}(x, \text{person}(x), \text{intend}(\text{go}(x))) & : p_\sigma \end{array}$$

- Narrow scope is undervivable

Cross-linguistic analysis

We have seen mainly from Japanese data:

- Mermaid Constructions as raising/equi
- Anaphoric control MC:
 - $(\uparrow \text{ PRED}) = \text{'N}_{\text{MC}} \langle (\text{SUBJ} | \text{OBL}_{\theta}), \text{COMP} \rangle^{25}$
 - $(\uparrow \text{ COMP SUBJ PRED}) = \text{'PRO'}$
- Functional control MC:
 - $(\uparrow \text{ PRED}) = \text{'N}_{\text{MC}} \langle \text{XCOMP} \rangle \text{SUBJ}$
 - $(\uparrow \text{ SUBJ}) = (\uparrow \text{ XCOMP SUBJ})$

How general are these formulations? Are they valid in other languages?
Arguments from here are ongoing work with sloppy points; comments are much appreciated.

²⁵ Whether SUBJ or OBL_θ is selected depends on each lexeme.

Dative case-marking in Tatar

Hypothesis 1: Control by OBL_{θ}

- *röxsät* subcategorizes for $\langle OBL_{\theta}, COMP \rangle$
- OBJ_{θ} controls ($\uparrow COMP$ SUBJ)

(54) Tatar (< Kipchak < Turkic; SOV)²⁶

Marat-qa [joqla-rya] *röxsät*

Marat-DAT sleep-INF permission

‘Marat is allowed to sleep.’

(55)
$$\left[\begin{array}{cc} \text{PRED} & \text{'permission}\langle OBL_{\theta}, COMP \rangle\text{' } \\ \text{OBL}_{\theta} & \left[\begin{array}{cc} \text{PRED} & \text{'Marat}_i\text{' } \\ \text{CASE} & \text{DAT} \end{array} \right] \\ & g \\ \text{COMP} & \left[\begin{array}{cc} \text{PRED} & \text{'sleep}\langle SUBJ \rangle\text{' } \\ \text{SUBJ} & \left[\text{PRED} \text{'PRO}_i\text{' } \right] \end{array} \right] \\ & h \\ f \end{array} \right]$$

²⁶ Chihiro Taguchi. “Mermaid Construction in Tatar”. In: *Proceedings of the 162nd Conference of the Linguistic Society of Japan*. 2021. 三 ▶

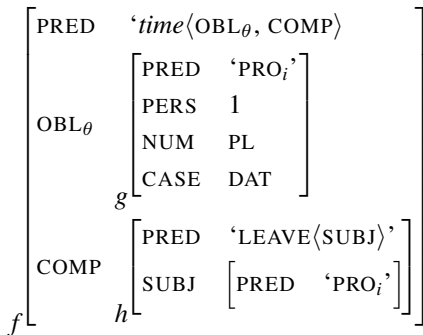
Dative case-marking in Russian

Seemingly working similarly

(56) Russian (< Slavic < Indo-European; SVO)

pora *nam* [uxodit']
time 1PL.DAT PF:leave:INF

'It is time for us to leave'



Prepositional phrase as OBL_{θ} in Welsh

In Welsh, the oblique argument is expressed by a prepositional phrase i

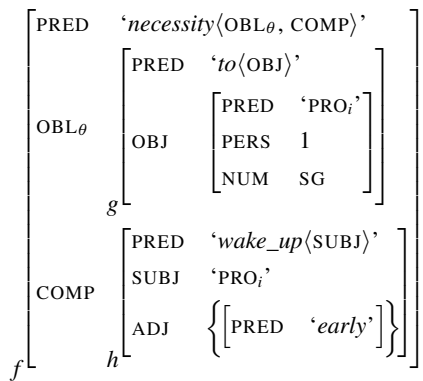
(57) Welsh (< Celtic < Indo-European; VSO)

Rhaid i fi [godi'n gynnar]

rhaid i fi [godi yn gynnar]

necessity to 1SG wake_up.INF in early

'I need to wake up early'



Genitive case-marking in Tatar

Hypothesis 2: Control by genitive SUBJ

- *isäp* subcategorizes for ⟨SUBJ, COMP⟩
- The possessive suffix on the MC noun is encoded as the value of PERS

(58) Tatar (< Kipchak < Turkic; SOV)²⁷

Marat-niŋ [joqla-rya] isäb-e

Marat-GEN sleep-INF thought-POSS.3

‘Marat is going to sleep.’

(59)
$$\left[\begin{array}{ll} \text{PRED} & \text{'thought'⟨SUBJ, COMP⟩} \\ \text{SUBJ} & \left[\begin{array}{ll} \text{PRED} & \text{'Marat}_i \text{' } \\ \text{CASE} & \text{GEN} \end{array} \right] \\ & g \\ \text{COMP} & \left[\begin{array}{ll} \text{PRED} & \text{'sleep'⟨SUBJ⟩} \\ \text{SUBJ} & \left[\begin{array}{ll} \text{PRED} & \text{'PRO}_i \text{' } \end{array} \right] \end{array} \right] \\ & h \\ \text{PERS} & 3 \end{array} \right]_f$$

²⁷ Chihiro Taguchi. “Mermaid Construction in Tatar”. In: *Proceedings of the 162nd Conference of the Linguistic Society of Japan*. 2021. 三 ▶

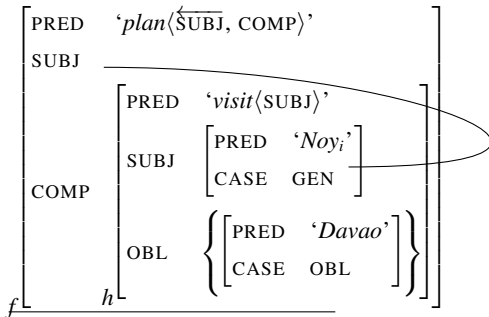
Genitive in embedded infinitival clause in Tagalog

How about this?

- (61) *plano*[=ng *b<um>isita ni=Noy sa=Davao bukas*]
 plan=LK visit<AF:INF> GEN=Noy OBL=Davao tomorrow
 ‘Noy plans to go to Davao tomorrow’

Hypothesis 3: Optional backward control raising in Tagalog

- Anaphoric controller in the matrix clause can be functionally controlled by the embedding SUBJ (backward control raising²⁹)



²⁹Susi Wurmbrand. “On forward and backward raising”. In: *Proceedings of the 32nd West Coast Conference on Formal Linguistics*. Somerville, MA, 2015.

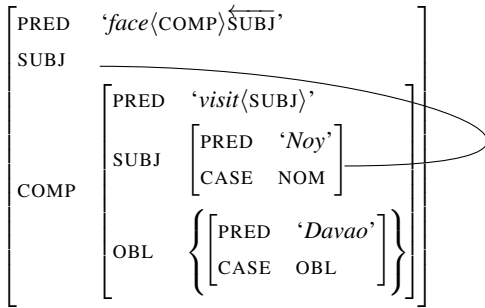
Finite COMP in Tagalog

mukha ‘face’ in Tagalog employs a finite COMP

- (62) *mukha*[=ng *bi-bisita* *si=Noy* *sa=Davao* *bukas*]
face=LK AF:CONT-visit **NOM**=Noy OBL=Davao tomorrow
'It seems that Noy will go to Davao tomorrow'

Hypothesis 4: Obligatory backward raising in Tagalog

- *mukha* ‘face’ obligatorily triggers backward raising³⁰



³⁰Susi Wurmbrand. “On forward and backward raising”. In: *Proceedings of the 32nd West Coast Conference on Formal Linguistics*. Somerville, MA, 2015.

Copy raising in Egyptian Arabic

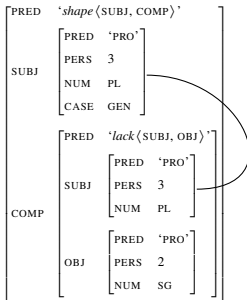
Egyptian Arabic MC (in a way) “repeats” the subject

(63) *šakl-uhum* [wahašū-k]
shape-POSS.3PL lack.PV.3PL-2SG.M

‘It seems like you missed them! (*lit.* their shape, you missed them)’

Hypothesis 5: Copy raising in Egyptian Arabic

- *šakl* in Egyptian Arabic causes copy raising³¹



³¹Shaimaa ElSadek and Louisa Sadler. In: *Proceedings of the LFG15 Conference*. CSLI Publications, 2015. »

Conclusion

This presentation has shown (from LFG22):

- Descriptive (theory-neutral) evidence for **biclausal** analysis of MC
- Syntactic evidence for **anaphoric/functional** control analysis
- **LFG's merits** to analyze MCs:
 - LFG readily allows for nominal predicates
 - COMP/XCOMP covering cross-linguistically diverse non-finite forms
 - Treating each module separately (syntax — (interface) — semantics)

New:

- Possible analyses for peculiar cases of MCs across languages
- Anaphoric/functional control analysis seems to hold
- Language-specific scrutiny is still much needed

Thank you!

Contact: ctaguchi@nd.edu

Website: <https://ctaguchi.github.io>