Harmonic Word Order Constraints are Not Created Equal: The Final-Over-Final Constraint as an Epiphenomenon

Joy Philip
j.philip@ucl.ac.uk
UCL

1 Introduction

We are indebted to the Final-over-Final Constraint (FOFC, Holmberg 2000; Biberauer, Holmberg & Roberts 2007, 2008a, 2008b, 2009, 2010; cf. Hawkins 1994; Julien 2002) for highlighting an unusual phenomenon in linearisation: the absence of certain disharmonic word orders.

FOFC is a descriptive observation generalising over one such absence, whereby a head-initial phrase cannot be dominated by a categorically non-distinct head-final phrase:

(1) The Final-over-Final Constraint

If $\alpha$ is a head-initial phrase and $\beta$ is a phrase immediately dominating $\alpha$, then $\beta$ must be head-initial. If $\alpha$ is a head-final phrase, and $\beta$ is a phrase immediately dominating $\alpha$, then $\beta$ can be head-initial or head-final, where:

(i) $\alpha$ and $\beta$ are in the same Extended Projection [categorically non-distinct, and $\alpha$P is a complement to $\beta$]$^1$.

(ii) $\alpha$P has not been A’-moved to Spec$\beta$P. (Biberauer, Holmberg & Roberts 2010:53, ex 1’’’’)

(2) Harmonic orders

<table>
<thead>
<tr>
<th>Initial-over-initial</th>
<th>Final-over-final</th>
<th>Initial-over-final</th>
<th>*Final-over-initial</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>$\alpha$</td>
<td>$\beta$</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>$\gamma$</td>
<td>$\alpha$</td>
<td>$\alpha$</td>
</tr>
</tbody>
</table>

However, there is some confusion as to whether FOFC is an invariant principle, or simply a non-absolute trend.

In this talk I make the following claims:

- FOFC only acts as a robust principle in regard to the distribution of subordinating complementisers.
- The distribution of subordinating complementisers can be derived independently of FOFC.
- Elsewhere, there is no absolute evidence for FOFC: it is at best a trend.

I will propose an alternative theory, whereby the presence or absence of disharmony is determined not by syntactic configuration, but by the presence or absence of certain semantic properties on a head. I will furthermore suggest that this alternative captures a wider range of data.

---

$^1$ Many thanks to my supervisor Ad Neeleman, for stimulating discussion, comments and encouragement, also to the audience at UCL PhD Day, 2nd December 2009. My thanks are also due to Hadja Habi Sali and Hamza Tidjani for Lagwan judgments. Further comments are welcome.

$^1$ Note that Biberauer et al.’s definition of Extended Projection differs from Grimshaw’s (1991, 2000).
2 Subordinating complementiser distribution

2.1 Complementiser distribution and FOFC

The most robust evidence for FOFC is found in certain left-right asymmetries in the distribution of subordinating complementisers (henceforth simply ‘complementiser’ or C).

FOFC makes the following predictions:

> **VO** languages allow only clause-initial Cs;

> **OV** languages allow both clause-initial Cs and clause-final Cs.

\( \text{(inter alia) } \text{Biberauer, Holmberg & Roberts 2007, 2008a, 2008b, 2009, 2010; Biberauer, Newton & Sheehan 2009) } \)

Where the CP is complement to a verb (and therefore categorially alike), FOFC predicts that a C-initial complement cannot be dominated by a final verb (Sheehan 2008):

> C-initial complement clauses must be postverbal.

These predictions are borne out in the attested distribution of Cs:

(3) **VO languages:**  **OV languages:**

<table>
<thead>
<tr>
<th>VO</th>
<th>OV</th>
<th></th>
<th>CVO</th>
</tr>
</thead>
<tbody>
<tr>
<td>*V[COV]</td>
<td>*[OVC]V</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>[CVO]</em></td>
<td><em>[COV]</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>[VOC]</em></td>
<td><em>V[OVC]</em></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


That is, while OV languages allow head-initial Cs, C-initial complement clauses cannot appear in canonical object position in such languages.

OV languages showing the V[COV] pattern include the Indo-Iranian languages Kudmali (or Kurmali), Maithili, Punjabi, Sindhi, Hindi-Urdu, Kashmiri, Pashto, Wakhi, Persian (or Farsi), Tajik and Zazaki, the West Germanic languages Afrikaans, Dutch and German, Latin (Italic), Hittite (Anatolian), Sorbian (Balto-Slavic), the Cushitic languages Iraqw and Somali, Neo-Aramaic (Semitic), the North Dravidian languages Brahuvi (or Brahvi) and Malto, the South Caucasian languages Georgian and Svan, the Atlantic-Congo languages Tunen and Sare\(^2\), the Malayo-Polynesian languages Gapapaiwa and Tawala, Anywa (or Anuak, Eastern Sudanic), Djapu (Pama-Nyungan), Mangarrayi (Gunwingguan), Mauka (or Mahou, Central-Southwestern Mande), Pari (Munduruku), Pima Bajo (Southern Uto-Aztecan), Teribe (Chibchan), Tsova-Tush (or Bats, Northeast Caucasian), Turkish (Turkic), Wappo (Yuki-Wappo) and Yaqui (Taracahitic).

(Dryer 1980, 2009; Bayer 2001; Cinque 2005; Davison 2007; Sheehan 2008; Biberauer, Holmberg & Roberts 2010)

---

\(^2\) Dryer (2009) also includes the Atlantic-Congo language Supyire in his list. However, an earlier example shows that in Supyire the rightwards extraposed clause is doubled by an overt pronoun in (preverbal) object position. Since the extraposed clause itself is not the complement of the verb, this example is irrelevant. Since Dryer does not give examples from all the other languages he lists, it is possible that some of these should also be excluded on the same grounds.
Harmonic Word Order Constraints are Not Created Equal.

5th Newcastle Postgraduate Conference in Linguistics, 23rd March 2010.

(4) a) Er hatte gewusst [dass er nicht lange leb-en würde].
German
he had known COMP he not long live-INF would
‘He had known that he would not live long.’

b) * Er hatte [dass er nicht lange leb-en würde] gewusst
he had COMP he not long live-INF would known

(Hawkins 1994:302, ex 5.43)

(5) a) An zan mi danat [ke an mard sangi partab kard].
Persian
that woman CONT knows COMP that man rock threw
‘The woman knows that the man threw a rock.’

b) * An zan mi [ke an mard sangi partab kard] danat
that woman CONT COMP that man rock threw knows

(Dryer 1980:130, exx 15-16)

(6) a) aapo hunen hia [ke hu hamut tutuʔuli].
Yaqui
he thus say COMP this woman pretty
‘He says that this woman is pretty.’

b) * aapo hunen [ke hu hamut tutuʔuli] hia
he thus COMP this woman pretty say

(p131, citing Lindenfeld 1973)

(7) a) Adam ban=a söyle-di-ø [ki Aysekitab=i oku-du-ø].
Turkish
man me=DAT tell-PST-3SG COMP Aysebook=ACC.DEF read-PST-3SG
‘A man told me that Ayse read the book.’

(ex 20)

b) * Adam ban=a [ki Aysekitab=i oku-du-ø] söyle-di-ø
man me=DAT COMP Aysebook=ACC.DEF read-PST-3SG tell-PST-3SG

(ex 21)

In Turkish, extraposition only occurs where there is a complementiser:

Turkish
man me=DAT Ayse=GEN book=ACC.DEF read-NOM-3SG.POSS=ACC tell-PST-3SG
‘The man told me that Ayse read the book/of Ayse’s reading the book.’

(Dryer 1980:131, ex 19)

everyone me=ACC.DEF Ankara=DAT go-PST consider-PROG
‘Everyone considers me to have gone to Ankara.’

(Özsoy 2001:217, ex 5a)

There is evidence that these **C-initial** complements are **base-generated in postverbal** position, since at least in Dutch, German, Hindi-Urdu, Persian and Turkish they are **not islands for extraction:**

(see Bennis 1987; Karimi 2001; Aghaei 2006; Biberauer, Newton & Sheehan 2009 and references cited in these works)

---

3 Abbreviations in glosses are as follows: ACC accusative; APPL applicative; ASP aspect; ASS assertive; AUX auxiliary; CL classifier; COMP complementiser; CONT continuous; DAT dative; DECL declarative; DEF definite; DEM demonstrative; DEP dependent; DU dual; FUT future; GEN genitive; IMMED immediate; IMP imperative; IMPF imperfective; INCOMPL incompletive; INDEF indefinite; INF infinitive; LOC locative; M masculine; n- n-word; N nominaliser; NEG negative; NOM nominative; OBJ object; PASS passive; PERF perfective; PL plural; POSS possessive; POT potential; PRES present; PROG progressive; PST past; Q interrogative; QUOT quotative; SBJ subject; SG singular; TOP topic

4 On the other hand, where extraposed clauses in these languages are doubled by a pronoun, and are therefore adjuncts, extraction is impossible (Karimi 2001; Aghaei 2006; Biberauer, Newton & Sheehan 2009 and references cited in these works).
Although it is therefore possible for OV languages to have base-generated postverbal complement, this option is not available where there is a final C: *V[OVC]

This is particularly striking in certain OV languages allowing both initial and final Cs: mainly Indo-Aryan languages with close geographical or historical contact with Dravidian, such as Assamese, Bengali (or Bangla), Oriya, Marathi, Gujarati and Nepali, and also the Turkic language Uzbek and the Atlantic-Congo language Vata (or Dida). (Bayer 1996, 1997, 1999, 2000, 2001; Cinque 2005; Davison 2007)

C-initial clauses are obligatorily postverbal; C-final clauses are obligatorily preverbal:

(11) a) chele-ta [or baba aS-be (bole)] Sune-che.
    boy-CL his father come-FUT.3 COMP hear-PST.3

b) chele-ta Sune-che [or baba aS-be (*?bole)].
    boy-CL hear-PST.3 his father come-FUT.3 COMP

‘The boy has heard that his father will come.’

(Bayer 1996:255, ex 9)

(12) a) chele-ta [(je) or baba aS-be] Sune-che
    boy-CL COMP his father come-FUT.3 hear-PST.3

b) chele-ta Sune-che [(je) or baba aS-be].
    boy-CL hear-PST.3 COMP his father come-FUT.3

‘The boy has heard that his father will come.’

(ex 11)

Evidence suggests that the C-initial clause is base-generated in postverbal position:

- The extraposed clause is c-commanded by the indirect object:

(13) ami [prottek-Ta chele-ke], bole-chi [je Ek-jon ta-ke, durga pujo-Y notun] jama kapoR de-be.
    I each-CL boy-OBJ say-PST.1 COMP one-CL he-OBJ Durga Puja-LOC new clothes give-FUT.3

‘I told [each boy], that someone will give him, new clothes at the festival of Durga Puja.’

(Bayer 2000:2, ex 5)

- The extraposed clause is not an island for extraction:

(14) kriSno mEleria-te, bhab-che [je ram ti, mara gE-che].
    Krishna malaria-LOC think-PERF.3 COMP Ram die go-PERF.3

‘Krishna thinks that Ram died of malaria.’

(Simpson & Bhattacharya 2000, ex 13)

---

5 It is debatable whether Vata genuinely has an initial complementiser. Koopman (1984) argues that only the final subordinator kā is a true complementiser.
2.2 Complementiser distribution and Head Proximity

The data in (3) can be summed up by the following two observations:

- **Final Cs** are only permitted in OV languages.
- **Subordinating Cs** must be base-generated such that they intervene linearly between their selecting verb and its complement clause.

Here I propose that both observations can be captured by the interaction of three independently-motivated **harmonic word order constraints**:

> **Head Proximity**


(15) **Principle of Head Proximity**

The head of a [subordinate] domain prefers to be contiguous with the head of its superordinate domain. (Rijkhoff 1986:5)

For example, Head Proximity accounts for the cross-linguistic tendency to avoid the placement of dependents of a noun between this noun and its selecting verb.


In (3), C, as head of the dependent clause, must be positioned such that it is adjacent to its selecting head, the matrix verb.


This refers to the preference for heads in a given language/domain to be uniformly head-initial/-final (cf. harmonic orders in (2)a) and b)).

As regards Cs, there should therefore be a preference for **initial Cs** in VO languages and for **final Cs** in OV languages.

> **CP-Final** (cf. Sentential NP Position Hierarchy, Dryer 1980; Language Independent Preferred Order of Constituents, Dik 1997)

Dryer (1980) showed that there is a cross-linguistic preference for **clausal arguments** to appear in **sentence-final** position.

Where these three constraints compete, **Head Proximity** takes precedence:
(16) **Harmonic Word Order Ranking**

In VO languages, the constraints do not compete, resulting in a single optimal order:

<table>
<thead>
<tr>
<th></th>
<th>Head Proximity</th>
<th>Head Uniformity</th>
<th>CP-Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>V[CVO]</td>
<td>![</td>
<td>![</td>
<td></td>
</tr>
<tr>
<td>*[VOC]V</td>
<td>![</td>
<td>![</td>
<td></td>
</tr>
<tr>
<td>*[CVO]V</td>
<td>![</td>
<td>![</td>
<td></td>
</tr>
<tr>
<td>*[V][VOC]</td>
<td>![</td>
<td>![</td>
<td></td>
</tr>
</tbody>
</table>

In OV languages, no single order obeys all three constraints.

In order to obey the dominant constraint, Head Proximity, either the Head Parameter or Final-over-elsewhere must be violated, resulting in two possible orders:

<table>
<thead>
<tr>
<th></th>
<th>Head Proximity</th>
<th>Head Uniformity</th>
<th>CP-Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>[OV]C</td>
<td>![</td>
<td>![</td>
<td></td>
</tr>
<tr>
<td>V[COV]</td>
<td>![</td>
<td>![</td>
<td></td>
</tr>
<tr>
<td>*[V][OV]</td>
<td>![</td>
<td>![</td>
<td></td>
</tr>
<tr>
<td>*[COV]V</td>
<td>![</td>
<td>![</td>
<td></td>
</tr>
</tbody>
</table>

3 FOFC elsewhere: trend rather than principle

We have seen that the distribution of Cs (in (3)) can be derived independently of FOFC, by the interaction of pre-existing word order constraints.

However, this result will be irrelevant if FOFC is required elsewhere.

For any category other than C, there is evidence suggesting that FOFC does not hold as an absolute principle: all four logically possible orders in (2), both harmonic and disharmonic, appear to be attested:

(Note that Dryer’s (1992) data only uses auxiliaries that are ‘specifically verbal’ and negative auxiliaries that ‘exhibit verbal properties’)

(19) a) Initial-over-initial: [P [N PossP]] = 134 languages (40%)
    b) Final-over-final: [[PossP N] P] = 177 languages (53%)
    c) Initial-over-initial: [P [PossP N]] = 14 languages (4%) (7% of N-final lgs)
    d) Final-over-initial: [[N PossP] P] = 11 languages (3%) (8% of N-initial lgs)

(Hawkins 2010:1, using data from Hawkins 1983)

(20) a) Initial-over-initial: [Aux [V O]] = 28 genera (39%)
    b) Final-over-final: [[O V] Aux] = 36 genera (51%)
    c) Initial-over-final: [Aux [O V]] = 3 genera (4%) (8% of OV genera)
    d) Final-over-initial: [[V O] Aux] = 4 genera (6%) (13% of VO genera)

(Data taken from Dryer 1992:100, table 28)

---

6 Dryer documents only the relative order of verb and object and of verb and auxiliary (i.e. not the relative order of object and auxiliary). However, since Julien (2002:235) states that the order V T/Asp O (where T/Asp is a free-standing morpheme) does not occur, it is safe to conclude that for the four genera exhibiting VO and VAux, the auxiliary follows the object.
(21) a) Initial-over-initial: \([\text{Neg} \ [V \ O]]\) = 13 genera (52%)
b) Final-over-final: \([\ [O \ V] \ \text{Neg}]\) = 8 genera (32%)
c) Initial-over-final: \([\text{Neg} \ [O \ V]]\) = 3 genera (12%) (27% of OV genera)
d) Final-over-initial: \([\ [V \ O] \ \text{Neg}]\) = 1 genus (4%) (7% of VO genera)

(Data taken from Dryer 1992:101, table 29)

(22) a) Initial-over-initial: \([Q \ [V \ O]]\) = 28 genera (30%)
b) Final-over-final: \([\ [O \ V] \ Q]\) = 32 genera (35%)
c) Initial-over-final: \([Q \ [O \ V]]\) = 13 genera (14%) (29% of OV genera)
d) Final-over-initial: \([\ [V \ O] \ Q]\) = 19 genera (21%) (40% of VO genera)

(Data taken from Dryer 1992:102, table 30)

Biberauer, Holmberg & Roberts (2007, 2008a, 2008b, 2009, 2010) claim that FOFC applies only to inflected auxiliaries (as opposed to uninflected T/Asp markers, which may not be heads).

(23) Ɂə di ɹa dɨˈrɔ, 1SG=see picture ASP
I’m looking at a picture.’

(Dryer 2008:20, ex 24, citing Henderson 1997:39)

(24) ca=do [m] Ɂə=khɔ phu má na]?
3=say COMP 3=FUT take what
‘What did he say he would take?’

(p21, ex 29, citing Henderson 1997:187)

However, the evidence suggests that the alleged ‘agreement inflection’ is in fact a proclitic pronoun, since it appears in complementary distribution with full NPs, in canonical subject position:

(25) fe ni dɔkhí tə-dó
trap catch barking.deer one-CL
‘the trap catches a barking deer’

(Dryer 2008:14, ex 2a, citing Henderson 1997:258)

(26) Ɂə=bɔdɔ [m] Ɂə=ce ḷe-nu ɜmɛ thɔ].
1SG-think COMP 1SG-book CL-that lost PERF
‘I thought that my book was lost’

(p21, ex 26, citing Henderson 1997:379)

Moreover, Matthew Dryer (p.c.) notes two Central Sudanic languages in his database with verbal auxiliaries with the order VOAux: Mbay and Ngambay.

> As regards negative and interrogative markers, Biberauer, Holmberg & Roberts (2010:§5) claim that they are syncategorematic: they are not involved in c-selection and are therefore outside the extended projection, hence outside the scope of FOFC.

However, in Ma’di (Central Sudanic), the marker of negation also encodes tense, a verbal feature, suggesting that it is indeed part of the extended projection of the verb:

(27) m’-åwi ɗòtɪ’kʊ, 1SG-open door NEG.NONPST
‘I won’t/am not opening the door/don’t open doors.’

(Blackings & Fabb 2003:14, ex 8)

---

\(^7\) Bailey et al (2010:13) and Biberauer, Holmberg and Roberts (2010:81) point out that some final interrogative particles may in fact be initial disjunctive elements with an elided second conjunct. The numbers for orders b) and d) in (22) may therefore be misleadingly high.
(28) m’-āwí ḏōt’ı̀kùrù.
1SG-open door NEG.PST
‘I did not open the door.’ (ex 7)

Notice that in the above examples, the lexical verb is unmarked for tense. However, in the absence of negation, the same past/nonpast distinction is marked on the lexical verb:

(29) ká ḍgbándà jà.
3SG cassava NONPST.eat
‘He is eating/eats cassava.’ (Blackings & Fabb 2003:13, ex 1)

Even where a final particle is uninflected, there may be evidence not only that the particle is a head dominating the verb, and that this head is involved in c-selection:

Lagwan (Chadic) exhibits the order TVONeg:

(30) Sà-dì gí r làsí yì bìyásí n sá.
FUT-3F go market tomorrow NEG
‘She won’t go to the market tomorrow.’

There is evidence for the projection of NegP, since the negative marker is required to license n-words:

(31) Bil=á shímá á ló *( sá).
man=LNK.M n-M 3MSG come NEG
‘Nobody came.’

There is evidence that final Neg dominates both initial V and T, since there is no true negative imperative; a surrogate (including T) is required (cf. Zanuttini 1994, 1996). This shows that Neg c-selects T:

(See also (30) above, where Neg follows (and therefore dominates) a TP-adjoined adverbial)

(32) Slà á!
push.IMP up
‘Get up!’

(33) a) * Slà á sá!
push.IMP up NEG
b) Sà-gí slá á sá!
FUT-2SG push up NEG
‘Don’t get up!’

Finally, there is evidence from the ‘Why not?’ test (Merchant 2001) that the negative marker in Lagwan is indeed a head, since it cannot adjoin to a phrasal category:

(34) a) * Àgé ghwànì sá?
because what NEG
b) * Sá ègé ghwànì?
NEG because what

4 (Dis)harmony and semantics

We have seen evidence suggesting then that disharmony is possible for any category except C, which always obeys the Harmonic Word Order Ranking in (16).
So, why do other categories not always obey this Harmonic Word Order Ranking? That is, why does disharmony arise?

I propose the following explanation:

- **Linearisation rules** pertaining to the **semantics** of a head can require it to appear in a **prominent position**, either initial or final.
- Where such rules **conflict** with, and **override**, the Harmonic Word Order Ranking in (16), **disharmony** arises.

For example, disharmony is relatively common for negative markers. The expression of negation is highly significant semantically. Cross-linguistically, there is a tendency to place negative markers in one of two prominent positions: initially, so that negation will be expressed as soon as possible (Jespersen 1917:4, 1924:297); or finally, the position reserved for new or significant information (Mazzon 2004:97). Where the choice of prominent position differs from the headedness of the verb, disharmony arises (cf. (21)c) and d)).

However, **subordinating C** is **impervious** to such **linearisation rules**:

- **Subordinating C** doesn’t contribute to the compositional semantics of its extended projection, but rather serves to **mark a relationship** between two extended projections; it is **semantically vacuous**.
- Disharmonic orders arise when linearisation rules require a head with specified semantics to appear in a certain (prominent) position; however, since C doesn’t have relevant semantics, the effects of **Head Proximity cannot be overridden** by such linearisation rules.

This makes two predictions:

> For any C that is not a subordinator, but rather **contributes to the compositional semantics**, **disharmony** should be possible.

While subordinating C is consistently harmonic (see section 2), interrogative markers and other discourse C-particles are frequent violators of both FOFC and its inverse (cf. (22), Julien 2002; Paul 2009, Biberauer, Holmberg & Roberts 2010):

(35) Ni yao kan zhe-ben shu ma?
    ‘Do you want to read this book?’

(36) a) Initial-over-initial: [N [REL TP]] = 56 languages (95%)
    b) Final-over-final: [[TP REL] N] = 3 languages (5%)
    c) *Initial-over-final: [N [TP REL]] = 0 languages (0%)
    d) *Final-over-initial: [[REL TP] N] = 0 languages (0%)

(37) a) Initial-over-initial: [N [LNK XP]] = 49 languages (66%)
    b) Final-over-final: [[XP LNK] N] = 25 languages (34%)

(Data taken from De Vries 2002:386-384, table 2)
c) *Initial-over-final: [N [XP LNK]] = 0 languages (0%)

d) *Final-over-initial: [[LNK XP] N] = 0 languages (0%)

(see Appendix B for languages and classification)

Here we find a complete absence of both disharmonic orders (cf. (2)). Note that neither absence falls under the descriptive/explanatory scope of FOFC.

Moreover, it may be necessary to violate FOFC in order to obey Head Proximity (note that linkers always form a constituent with the dependent of the noun, Philip 2009):

(38) [ zuo\textit{tian} chi yurou \textit{de}] ren
    \begin{tabular}{l}
    \textit{yesterday eat fish} \text{LNK person} \\
    \text{‘The people who ate fish yesterday’}
    \end{tabular}

\hspace{1cm} \text{(Paul 2009:4, ex 8a)}

5 Summary and conclusion

- Subordinating complementisers (and other semantically vacuous subordinators) always obey optimal word order principles, governed by the interaction between Head Proximity, Head Uniformity and CP-Final, with Head Proximity taking precedence. This allows us to capture a wider range of absent disharmonic word orders than FOFC.
- This is because linearisation rules pertaining to the semantics of a head cannot apply to heads lacking semantics.
- Elsewhere, such linearisation rules are permitted to override harmonic word order principles, giving rise to disharmony, including FOFC violations.

Appendix A: Superficial counterexamples

The literature cites various (superficial) counterexamples to the observations in (3).

These fall under one or both of two categories:

- The complement clause is (embedded under a) nominal, therefore not a direct complement of V.
- The postverbal complement clause is not base-generated there as a complement of the verb, but appears in an adjoined position (either by movement, or it is an adjunct associated with a null nominal).

> Harar Oromo (or Galla, Cushitic), allegedly [COV]V:

(39) inn\textit{íí} [ akka deem-u] good’\textDash ám-\textit{é}.

\begin{tabular}{l}
\textit{he that go-DEP order-PASS-PST} \\
\text{‘He was ordered to go.’}
\end{tabular}

\hspace{1cm} \text{(Owens 1985:145, ex 49)}

Owens (1985) refers to the embedded clause headed by akka as a ‘\textit{noun clause’}. Where an akka-clause expresses purpose, it is optionally marked with the dative case:

(40) [ akka na \textit{árk-anfí-f}]d’uf-an.

\begin{tabular}{l}
\textit{that me see-PL-DAT came-PL} \\
\text{‘They came to see me.’}
\end{tabular}

\hspace{1cm} \text{(Owens 1985:146, ex 54)}

Elsewhere, the lack of case-marking is expected, since ordinary direct objects appear in the morphologically unmarked absolutive case.
Moreover, the semantics of akka suggest that it is not a subordinating complementiser: ‘according as, just as, like, how, manner, way, (in order) to/that, (the fact) that’

(Hodson & Walker 1922; Owens 1985)

> Lakhota (or Lakota, Siouan), allegedly V[OVC] (Dryer 1980):

(41) [Tohá slolyáya he] [wakpála ektá ohį̄paye kį].

   ‘When did you know that he fell in the creek?’

   (Rood 1973:73, ex 8)

kį, which Rood (1973) terms a nominaliser, is identical to the definite determiner, which also appears in final position.

It appears in complementary distribution with an indefinite marker:

(42) a) [Tuwá omákiyį̄ kte kį̄ slolwáye.]  Lakhota

   who help.me POT the I.know

   ‘I know who will help me.’

b) [Tuwá omákiyį̄ čha slolwáye.]

   who help.me POT INDEF I.know

   ‘I know somebody will help me.’

   (Rood 1973:72, ex 6)

For verbs such as ‘say’, ‘want’, ‘tell’, ‘be able’, ‘be supposed to’ a bare clause appears as complement. kį̄ (definite) or čha (indefinite) are only required with certain verbs. This suggests that the latter set of verbs select nominal, rather than clausal, complements.

In any case, the extraposition in (41) is optional.

Moreover, the position of the complement following the interrogative marker he suggests that it is adjoined externally to the matrix clause, rather than base-generated as a complement.

> Dhivehi (or Maldivian, Indo-Iranian), allegedly V[OVC] (Cinque 2005):

The complement clause is marked by dative or locative case:

(43) ahannaš lafā kurevenī [hama jessēne kam-aš].

   me.DAT guess do.IN.PREFOC even touch.FUT thing.DAT

   ‘I guess (the event) that it will even out.’

   (Cain & Gair 2000:48, ex 163)

> Ngiti (Central Sudanic), allegedly V[OVC] (Cinque 2005):

The complement clause is embedded under a noun:

(44) k=ùnì [ma m-i-rā dhu].

   3SG=know.PERF.PRES I 1-AUX-come.N thing

   ‘He knows (the fact) that I am coming.’

   (Kutsch Lojenga 1994:395)

> Khoekhoe (or Nama, Khoe), allegedly V[OVC] (Dryer 2009):

The complement clause is marked by objective case:

(45) tsìi //ũi-p-ã-kxînì ke kē mĩ́-pa [!ũi-kxînì ta !xáis=ã].

   and 3-MSG-OBJ-1M.DU.SBJ DECL PST say-APPL go-1M.DU.SBJ COMP=OBJ

   ‘And we told him that we were going.’

   (Hagman 1977:138, gloss Güldemann 2006, ex 20)
### Appendix B: Languages with linkers in the noun phrase:

<table>
<thead>
<tr>
<th>Classification</th>
<th>No. of lgs. in sample</th>
<th>Language</th>
<th>Position of linker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afro-Asiatic</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chadic</td>
<td>(9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biu-Mandara</td>
<td>(8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>(1)</td>
<td>Gude</td>
<td>Postnominal</td>
</tr>
<tr>
<td>B</td>
<td>(7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kotoko-Yedina</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kotoko</td>
<td>(5)</td>
<td>Afade</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Goulfey</td>
<td></td>
<td></td>
<td>Postnominal</td>
</tr>
<tr>
<td>Lagwan</td>
<td></td>
<td></td>
<td>Postnominal</td>
</tr>
<tr>
<td>Mpade</td>
<td></td>
<td></td>
<td>Postnominal</td>
</tr>
<tr>
<td>Mser</td>
<td></td>
<td></td>
<td>Postnominal</td>
</tr>
<tr>
<td>Biu-Mandara</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zina</td>
<td>(2)</td>
<td>Mazera</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Zina</td>
<td></td>
<td></td>
<td>Postnominal</td>
</tr>
<tr>
<td>Chadic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West Chadic</td>
<td>(1)</td>
<td>Nyam</td>
<td>Postnominal</td>
</tr>
<tr>
<td>East Cushitic</td>
<td>(1)</td>
<td>Dasenech</td>
<td>Prenominal</td>
</tr>
<tr>
<td>Austronesian</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malayo-Polynesian</td>
<td>(6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central Philippine</td>
<td>(1)</td>
<td>Tagalog</td>
<td>Both</td>
</tr>
<tr>
<td>Oceanic</td>
<td>(5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central-Eastern Oceanic</td>
<td>(2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central Pacific</td>
<td>(1)</td>
<td>Rotuman</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Micronesian</td>
<td>(1)</td>
<td>Kiribati</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Meso-Melanesian</td>
<td>(1)</td>
<td>Bali-Vitu</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Polynesian</td>
<td>(1)</td>
<td>Samoan</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Southern Oceanic</td>
<td>(1)</td>
<td>Malo</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Nuclear Malayo-Polynesian</td>
<td>(1)</td>
<td>Palauan</td>
<td>Both</td>
</tr>
<tr>
<td>Tsouic</td>
<td>(1)</td>
<td>Tsou</td>
<td>Prenominal</td>
</tr>
<tr>
<td>Basque</td>
<td>1</td>
<td>Basque</td>
<td>Prenominal</td>
</tr>
<tr>
<td>Creole languages</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dutch Creole</td>
<td>(1)</td>
<td>Berbice Dutch Creole</td>
<td>Postnominal</td>
</tr>
<tr>
<td>English Creole</td>
<td>(1)</td>
<td>Tok Pisin</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Indo-European</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Albanian</td>
<td>(1)</td>
<td>Albanian</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Indo-Iranian</td>
<td>(11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indo-Aryan</td>
<td>(2)</td>
<td>Hindi</td>
<td>Prenominal</td>
</tr>
<tr>
<td>Southern Iranian</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Western Iranian</td>
<td>(9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northwestern Iranian</td>
<td>(7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caspian</td>
<td>(2)</td>
<td>Gilaki</td>
<td>Both</td>
</tr>
<tr>
<td>Mazendarani</td>
<td></td>
<td></td>
<td>Both</td>
</tr>
<tr>
<td>Kurdish</td>
<td>(5)</td>
<td>Balochi</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Southern Iranian</td>
<td>(2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southwestern Iranian</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Western</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italic</td>
<td>(3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Romance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>East Romance</td>
<td>(1)</td>
<td>Romanian</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Italo-Western</td>
<td>(2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italo-Dalmation</td>
<td>(1)</td>
<td>Italian</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Western</td>
<td>(1)</td>
<td>French</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Language Group</td>
<td>Number</td>
<td>Language</td>
<td>Word Order</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>--------</td>
<td>-------------------</td>
<td>------------</td>
</tr>
<tr>
<td>West Germanic</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anglo-Frisian</td>
<td>1</td>
<td>English</td>
<td>Postnominal</td>
</tr>
<tr>
<td>High German</td>
<td>1</td>
<td>German</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Low Franconian</td>
<td>1</td>
<td>Dutch</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Japonic</td>
<td>1</td>
<td>Japanese</td>
<td>Prenominal</td>
</tr>
<tr>
<td>Korean</td>
<td>1</td>
<td>Korean</td>
<td>Prenominal</td>
</tr>
<tr>
<td>Kwadi-Khoe</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Khoe</td>
<td>1</td>
<td>Khoekhoe</td>
<td>Prenominal</td>
</tr>
<tr>
<td>Mayan</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cholan-Tzeltalan</td>
<td>1</td>
<td>Tzeltal</td>
<td>Prenominal</td>
</tr>
<tr>
<td>Niger-Congo</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atlantic-Congo</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benue-Congo</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bantu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zone D</td>
<td>1</td>
<td>Kilega</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Zone E</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kikuyu-Kamba</td>
<td></td>
<td>Gikuyu</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Zone G</td>
<td>1</td>
<td>Swahili</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Zone J</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haya-Jita</td>
<td>1</td>
<td>Haya</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Konzo</td>
<td>1</td>
<td>Kinande</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Zone N</td>
<td>1</td>
<td>Chichewa</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Zone S</td>
<td>1</td>
<td>Shona</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Senegal-Guinea</td>
<td>1</td>
<td>Wolof</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Western Mande</td>
<td>1</td>
<td>Bambara</td>
<td>Prenominal</td>
</tr>
<tr>
<td>Nilo-Saharan</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central Sudanic</td>
<td>1</td>
<td>Lendu</td>
<td>Prenominal</td>
</tr>
<tr>
<td>East Sudanic</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Western Nilotic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southern Luo</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luo-Acholi</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Songhay</td>
<td>1</td>
<td>Koyra Chiini</td>
<td>Prenominal</td>
</tr>
<tr>
<td>Penutian</td>
<td>1</td>
<td>Tsimshian</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Sino-Tibetan</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sinitic</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td></td>
<td>Cantonese</td>
<td>Prenominal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mandarin</td>
<td>Prenominal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Taiwanese</td>
<td>Prenominal</td>
</tr>
<tr>
<td>Tibeto-Burman</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Himalayish</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mahakiranti</td>
<td>1</td>
<td>Newari</td>
<td>Prenominal</td>
</tr>
<tr>
<td>Tibeto-Kanauri</td>
<td>1</td>
<td>Byansi</td>
<td>Prenominal</td>
</tr>
<tr>
<td>Lolo-Burmese</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burmish</td>
<td>1</td>
<td>Burmese</td>
<td>Prenominal</td>
</tr>
<tr>
<td>Loloish</td>
<td>1</td>
<td>Lahu</td>
<td>Prenominal</td>
</tr>
<tr>
<td>Northeast Tibeto-Burman</td>
<td>1</td>
<td>Bai</td>
<td>Prenominal</td>
</tr>
<tr>
<td>Tai-Kadai</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tai</td>
<td></td>
<td>Thai</td>
<td>Postnominal</td>
</tr>
<tr>
<td>Trans-New Guinea</td>
<td>1</td>
<td>Amele</td>
<td>Prenominal</td>
</tr>
<tr>
<td>Madang</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References

Biberauer, Theresa, Anders Holmberg & Ian Roberts. 2010. ‘A Syntactic Universal and its Consequences.’ Ms, University of Cambridge, Stellenbosch University & Newcastle University.


