Understanding morphosyntactic variation in a temporally and spatially representative Warlpiri corpus

A preliminary report on word order in clauses

Maria Vollmer (Freiburg, ANU, CoEDL)
I gratefully acknowledge the Warlpiri people who provided the material used in this presentation. I acknowledge that the Warlpiri data was recorded on Warlpiri and Gurindji land and I pay my respects to their elders past, present and emerging.
Content and aim of this poster

- Part of a PhD thesis (2nd year) on language contact and change in word order, ergative case marking and parts-of-speech
- Content: Word order variation in the finite verbal clause in newer materials from Lajamanu
- Preliminary, qualitative observations
Research background

- Initial description of Warlpiri as having no basic word order
  (such as Hale 1983; Hale 2002: 117)

- Then some (preliminary) evidence of preferred word order with pragmatically motivated variations
  (such as Hale 1992: 76; Hale et al. 1995: 1431; Swartz 1991: 55; Simpson 2007; Simpson & Mushin 2008)

Corpus examples of flexible word order:

SVO
(1) \textit{wati-ngki luwa-rnu puluku} \\
    man-ERG shoot-PST bullock \\
    `The man shot the bullock.' (wa32-2-8-9)

SOV
(2) \textit{kuuku-ngu ka jarntu ma-ni} \\
    monster-ERG AUX.PRS dog get-NPST \\
    `The monster gets the dog.' (wa32-1)

Corpus source: O’Shannessy
Research background

- Initial description of Warlpiri as having no basic word order
  (such as Hale 1983; Hale 2002: 117)

- Then some (preliminary) evidence of preferred word order with pragmatically motivated variations
  (such as Hale 1992: 76; Hale et al. 1995: 1431; Swartz 1991: 55; Simpson 2007; Simpson & Mushin 2008)

Corpus examples of flexible word order:

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SV</td>
<td>kamina-jarra ka =pala karri-mi</td>
<td></td>
</tr>
<tr>
<td></td>
<td>girl-DL AUXPRS AUX3DL.SBJ stand-NPST</td>
<td></td>
</tr>
<tr>
<td></td>
<td>`The two girls are standing (there).' (wa32-1)</td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>VS</td>
<td>nyina-mi ka malju [...]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sit-NPST AUXPRS boy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>`The boy sits [...] by the fire].' (wa32-1)</td>
<td></td>
</tr>
</tbody>
</table>

Corpus source: O’Shannessy
Research background

- Initial description of Warlpiri as having no basic word order
  (such as Hale 1983; Hale 2002: 117)

- Then some (preliminary) evidence of preferred word order with pragmatically motivated variations
  (such as Hale 1992: 76; Hale et al. 1995: 1431; Swartz 1991: 55; Simpson 2007; Simpson & Mushin 2008)

But auxiliary obligatory in second position:

SV
(3) kamina-jarra ka =pala karri-mi
   girl-DL AUX.PRS AUX.3SBJ stand-NPST
   `The two girls are standing (there).’ (wa32-1)

VS
(4) nyina-mi ka malju [...]
   sit-NPST AUX.PRS boy
   `The boy sits [...] by the fire’ (wa32-1)

Corpus source: O’Shanessy
Research background

But none of that was based on representative corpus across time and space

1. Building representative corpus to control for areal variation and language change

2. Annotation with GRAID and reference tracking (extensive, consistent)

-> Allows for a multifactorial analysis (i.e. with decision trees)
## Building a representative corpus

<table>
<thead>
<tr>
<th>Place of recording</th>
<th>Time of recording</th>
<th>Amount of clause units</th>
<th>Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lajamanu</td>
<td>50s to 70s (?)</td>
<td>250</td>
<td>?</td>
</tr>
<tr>
<td>Lajamanu</td>
<td>after 2000</td>
<td>253</td>
<td>Carmel O'Shanessy, Nelson 2009</td>
</tr>
<tr>
<td>Yuendumu</td>
<td>50s to 70s</td>
<td>250</td>
<td>Ken Hale</td>
</tr>
<tr>
<td>Yuendumu</td>
<td>after 2000</td>
<td>263</td>
<td>Carmel O'Shanessy, Daniels 2009</td>
</tr>
<tr>
<td>Willowra</td>
<td>50s to 70s</td>
<td>257</td>
<td>Ken Hale</td>
</tr>
<tr>
<td>Willowra</td>
<td>after 2000</td>
<td>268</td>
<td>Carmel O'Shanessy, Morton 2009a/b/c/d/e, Presley 2009</td>
</tr>
</tbody>
</table>
## Data used for this poster

<table>
<thead>
<tr>
<th>Place of recording</th>
<th>Time of recording</th>
<th>Amount of clause units</th>
<th>Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lajamanu</td>
<td>50s to 70s (?)</td>
<td>250</td>
<td>?</td>
</tr>
<tr>
<td>Lajamanu</td>
<td>after 2000</td>
<td>253</td>
<td><a href="#">Carmel O'Shanessy, Nelson 2009</a></td>
</tr>
<tr>
<td>Yuendumu</td>
<td>50s to 70s</td>
<td>250</td>
<td><a href="#">Ken Hale</a></td>
</tr>
<tr>
<td>Yuendumu</td>
<td>after 2000</td>
<td>263</td>
<td><a href="#">Carmel O'Shanessy, Daniels 2009</a></td>
</tr>
<tr>
<td>Willowra</td>
<td>50s to 70s</td>
<td>257</td>
<td><a href="#">Ken Hale</a></td>
</tr>
<tr>
<td>Willowra</td>
<td>after 2000</td>
<td>268</td>
<td><a href="#">Carmel O'Shanessy, Morton 2009a/b/c/d/e, Presley 2009</a></td>
</tr>
</tbody>
</table>
Building a representative corpus

- Coded on different levels with GRAID, RefIND, ISNRef, and “own“ coding (Haig & Schnell 2014, Schiborr et al. 2018)
- Complex coding; coding = analysis in itself

### (5) `The horse is lying on the ground.’ (wa-32-2-11)

```plaintext
## nantuwu ka nguna walya-ngka
## horse AUX.PRS lie.PRS ground-LOC
## np.an:s aux v:pred np:l
009 0027

`The horse is lying on the ground.’ (wa-32-2-11)
```

For detailed info on the GRAID coding, see Vollmer (to appear); corpus source: O’Shanessy.
The horse is lying on the ground.

Distinct number for each referent

Newly introduced referents are also tagged with new, bridging (i.e. man was part of already mentioned group etc.), or unused

(5) # nantuwu ka nguna walya-ngka
## horse AUX.PRS lie.PRS ground-LOC
## np.an:s aux v:pred np:l
009 0027
`The horse is lying on the ground.' (wa-32-2-11)

- Coded on different levels with GRAID, RefIND, ISNRef, and “own“ coding  
  (Haig & Schnell 2014, Schiborr et al. 2018)

- Complex coding; coding = analysis in itself

For detailed info on the GRAID coding, see Vollmer (to appear); corpus source: O’Shannessy.
• Coded on different levels with GRAID, RefInd, ISNRef, and “own” coding (Haig & Schnell 2014, Schiborr et al. 2018)

• Complex coding; coding = analysis in itself

\(\text{(5)}\)  
\[
\begin{array}{llll}
\# nantuwe & ka & nguna & walya-ngka \\
\# horse & AUX.PRS & lie.PRS & ground-LOC \\
\# np.an:s & aux & v:pred & np:l \\
009 & & 0027 \\
\end{array}
\]

`The horse is lying on the ground.' (wa-32-2-11)
The horse is lying on the ground.

Form . animacy : function

np = noun; pro = pronoun; 0 = zero
h = human; an = animate; ø = non-human
s = intransitive subject; a = transitive subject; p = object; g = goal; l = location; obl = other obliques

Coded on different levels with GRAID, RefIND, ISNRef, and “own” coding (Haig & Schnell 2014, Schiborr et al. 2018)

Complex coding; coding = analysis in itself

For detailed info on the GRAID coding, see Vollmer (to appear); corpus source: O’Shannessy.
The horse is lying on the ground.' (wa-32-2-11)

Form . animacy : function
np = noun
ø = non-human
l = location

• Coded on different levels with GRAID, RefIND, ISNRef, and “own“ coding (Haig & Schnell 2014, Schiborr et al. 2018)

• Complex coding; coding = analysis in itself

(5)  # nantuwu  ka  nguna  walya-ngka
    # horse  AUX.PRS  lie.PRS  ground-LOC
    np.an:s  aux  v:pred  np:l
    009  0027

`The horse is lying on the ground.’ (wa-32-2-11)
The horse is lying on the ground.' (wa-32-2-11)

- Coded on different levels with GRAID, RefIND, ISNRef, and “own“ coding (Haig & Schnell 2014, Schiborr et al. 2018)

- Complex coding; coding = analysis in itself

For detailed info on the GRAID coding, see Vollmer (to appear); corpus source: O’Shannessy.
RefIND tells us:
- Givenness/Newness of referents
- How referents are introduced
- Overall frequency of referents
- Topicality of referents (indirectly)
- Last and subsequent mentions of referent

GRAID tells us:
- Word order
- Animacy (human, animate, non-human)
- Form (noun, zero, pronoun)
- Function (subject/object/oblique)

For detailed info on the GRAID coding, see Vollmer (to appear); corpus source: O’Shannessy.

(5)  ## nantuwu  ka  nguna  walya-ngka
    ## horse  AUX.PRS  lie.PRS  ground-LOC
    ## np.an:s  aux  v:pred  np:l
009  0027

`The horse is lying on the ground.’ (wa-32-2-11)
Building a representative corpus

Not all relevant information is captured 😢

Prosody: Segmentation into major (//) and minor (/) intonation units

(6)  

\# nantuwu ka nguna walya-ngka //  
\# horse AUX.PRS lie.PRS ground-LOC  
\# np.an:s aux v:pred np:l  
009 0027

`The horse is lying on the ground.’ (wa-32-2-11)

I appreciate any feedback on how I could do more detailed prosodic coding — there will be audio examples in the preliminary results section!
Initial results
What “triggers“ VS order?
(7) \textit{nyina-mi ka malju}\textsuperscript{QR}
noina-mi AUX boy
`Sits the boy, [...] [by the fire].'
(wa32-1-084-086; O’Shannessy)

Possible triggers of VS order
1. Contrastive focus on the verb
2. Switch reference
3. New referent (bridging)

Discourse context
\textit{kamina jinta-ngku ka // jarntu manu yard-jangka // \textit{nyinami ka malju / kurdu witapardu / warluwana / kamina-jarra ka =pala karri-mi / manu jarntu ka nguna-mi warlu-wana //}}

one girl gets the dog from the yard. \textbf{sits the boy, the little child, by the fire.} the two girls are standing (there) and the dog lies by the fire
(8) \( \text{rarralykaji-rla} =\!\text{lu} \quad \text{ya-nu} \quad \text{wirlinyi} \quad // \quad \text{wati} \quad \text{mankurrpa} \ / \ \\
\text{car-LOC} \quad \text{AUX.3PL.SBJ} \quad \text{go-PST} \quad \text{daytrip} \quad \text{man} \quad \text{a.few} \)

`Went hunting in the car a few men.’ (wa32-3-004-005; O’Shan Nessy)

Possible triggers of VS order:
1. Afterthought (= different IU?)
2. Discourse organisation: speaker is unsure?
3. New referents?
What factors play into OV versus VO positioning?
Caveat

• **Intonation Unit breaks** more common in VO word order (around 50% of all instances of VO)

• Frequently **afterthoughts**

→ Reason for slightly higher number of VO?
Open question and sidenote: Afterthoughts are difficult to annotate because of the flexible word order, but would a consistent annotation be possible on the basis of (a) specific prosodic contour(s)/intonation unit breaks?

wa32-3-014-015; O’Shannessy
wa32-2-008-009; O’Shannessy; unsure if afterthought?
wa29-3-039-40; O’Shannessy

j15-061-062; Nelson 2009
wa32-2-013-014; O’Shannessy
wa32-2-006-007; O’Shannessy; O is relativised in same IU
Focus on verb makes VO likely

• Verb before auxiliary in around a third of VO

• Rarely contrastive focus, but more focus/prominence/a sequence of events(?)

(11) nya-nyi ka =lu // kanta
    see-PST AUX.PRS AUX.3PL.SBJ bush.coconut
`(They) saw bush coconuts.'
(wa32-3-014-5; O'Shannessy)
Given objects make VO more likely

*In finite verbal clauses

Low numbers: Preliminary
Given objects make **VO** more likely

(12) \[\text{wirnpa-ngku} \quad \text{ka} \quad \text{luwa-rni} \quad \text{wati}\]
\text{lightning-ERG} \quad \text{AUX.PRS} \quad \text{hit-NPST} \quad \text{man}

`Lightning hits the man.’
(wa32-3-052; O'Shannessy)

(13) \[\text{nganimpa} \quad \text{laju} \quad \text{paji-rni}\]
\text{we} \quad \text{edible.grub} \quad \text{pick-NPST}

`[...] we pick edible grubs [...]’
(j15-034; Nelson 2009)
English loan words make VO more likely

- English loans more frequent in VO (preliminary)

- English (SVO) influence?

(14) \textit{wati-ngi manu kurdu-ngu ka man-ERG and child-ERG AUX.PRS =pala ma-ni fenci AUX.3DL.SBJ get-NPST fence} \\
\textit{`The man and the child get the fence.'} (wa32-2-040; O'Shannessy)
English loan words make **VO** more likely
Semantically weak verbs make OV more likely

Clauses with semantically weak verbs *ma-ni* (take, get), *yirra-rni* (put, make, do), *marda-rni* (have, hold) occur in OV around 66% of the time, and in VO around 33%.

(15) `kuuku-ngu ka jarntu ma-ni`
man-ERG AUX.PRS dog get-NPST
`The monster gets the dog.’
(wa32-1-049; O’Shannessy)

(16) `jumu yirra-rnu`
well dig-PST
`dug a native well’ (j15-022; Nelson 2009)
What factors play into OV versus VO order?

- Focus on verb (before auxiliary) makes VO likely
- Given objects make VO more likely
- English loan words make VO more likely
- Semantically weak verbs make OV more likely

Low numbers: Preliminary
Conclusion, Outlook and other things
Initial results

- SV is preferred to VS
- No clear preference for OV vs. VO
- Factors:
  - Focus on verb
  - Information status
  - Afterthought
  - English loan words
  - Semantically weak verbs
Limitations

- Preliminary observations
- Not enough data points for statistical significance
- No detailed quantitative analysis of all factors
- Very first, initial part of a bigger and more extensive PhD thesis, including areal and temporal variation
Outlook

- Control for areal variation and language change
- More data = statistical significance
- More complex quantitative analysis
- Figure out weight of each factor, e.g. with decision trees
- Bigger-picture PhD thesis: Interplay with ergative case marking and word order within nominal expressions while controlling for areal variation and language contact/change
1. I appreciate any feedback on anything!

2. Would it be useful to code *prosody* in more detail, and do you have suggestions for consistent coding?

3. Do intonation unit breaks always signal afterthoughts, and should I try to make a distinction between VO in a single versus a separate intonation unit?

4. Might it be worth coding semantically weak verbs given my preliminary results?

5. Do you know of any Warlpiri texts collected between 1950 and 1970 in Lajamanu?

6. Am I missing any important factors for word order (that could feasibly be annotated/included in the analysis)?


I want to thank the Warlpiri speaking communities who provided the material used in this presentation. I acknowledge that the Warlpiri data was recorded on Warlpiri and Gurindji land and I pay my respects to their elders past, present and emerging.

Many thanks to Jane Simpson and David Nash who have helped extensively with the coding of the data and with finding archived texts for the corpus. This project would not be possible without you.

I am incredibly grateful to Carmel O'Shannessy who was kind enough to share her recorded data with me and allow me to use it for analysis. Thank you so much!

Thank you to the Multi-CAST team Geoffrey Haig, Stefan Schnell and Nils Schiborr who are always happy to give feedback on and discuss the coding of constructions with GRAID.

Sarah Stolle has helped tremendously by coding part of the corpus with GRAID, RefIND, and ISNRef; and by providing the data necessary for inter-rater reliability testing.

Thanks to James Gray who has provided useful feedback and a safe space for any kind of brainstorming, no matter how spontaneous.

My PhD thesis is supported by funding from the Deutsche Forschungsgemeinschaft project number 406074683, and the ARC Centre of Excellence for the Dynamics of Language project ID CE140100041.
Thank you!

Maria Vollmer (Freiburg, ANU, CoEDL)

@vollmer_maria