

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) *wati-ngki luwa-rnu puluku*
man-ERG shoot-PST bullock
'The man shot the bullock.' (wa32-2-8-9)

SOV

- (2) *kuuku-ngu ka jarntu ma-ni*
monster-ERG AUX.PRS dog get-NPST
'The monster gets the dog.' (wa32-1)

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:

SV

(3) *kamina-jarra ka =pala karri-mi*
girl-DL AUX.PRS AUX.3DL.SBJ 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)

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.3DL.SBJ 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)

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

Place of recording	Time of recording	Amount of clause units	Source(s)
Lajamanu	50s to 70s (?)	250	?
Lajamanu	after 2000	253	Carmel O'Shannessy, Nelson 2009
Yuendumu	50s to 70s	250	Ken Hale
Yuendumu	after 2000	263	Carmel O'Shannessy, Daniels 2009
Willowra	50s to 70s	257	Ken Hale
Willowra	after 2000	268	Carmel O'Shannessy, Morton 2009a/b/c/d/e, Presley 2009

Data used for this poster

Place of recording	Time of recording	Amount of clause units	Source(s)
Lajamanu	50s to 70s (?)	250	?
Lajamanu	after 2000	253	Carmel O'Shannessy, Nelson 2009
Yuendumu	50s to 70s	250	Ken Hale
Yuendumu	after 2000	263	Carmel O'Shannessy, Daniels 2009
Willowra	50s to 70s	257	Ken Hale
Willowra	after 2000	268	Carmel O'Shannessy, Morton 2009a/b/c/d/e, Presley 2009

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)	##	<i>nantuwu</i>	<i>ka</i>	<i>nguna</i>	<i>walya-ngka</i>
	##	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

Distinct number for each referent

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

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

* The referents in this example are not newly introduced; *new* and *bridging* have been added only for demonstration purposes.

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

Clause boundary

(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)

Form . animacy : function

np = noun; pro = pronoun; 0
= zero

h = human; an = animate; ∅ =
inanimate

s = intransitive subject; a =
transitive subject; p = object;
g = goal; l = location; obl =
other obliques

- | | | | | | |
|-----|----|----------------|-----------|--------------|-------------------|
| (5) | ## | <i>nantuwu</i> | <i>ka</i> | <i>nguna</i> | <i>walya-ngka</i> |
| | ## | 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)

(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)

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)

Building a representative corpus

Not all relevant information is captured 😓 (5) ## *nantuwu ka nguna walya-ngka //*

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

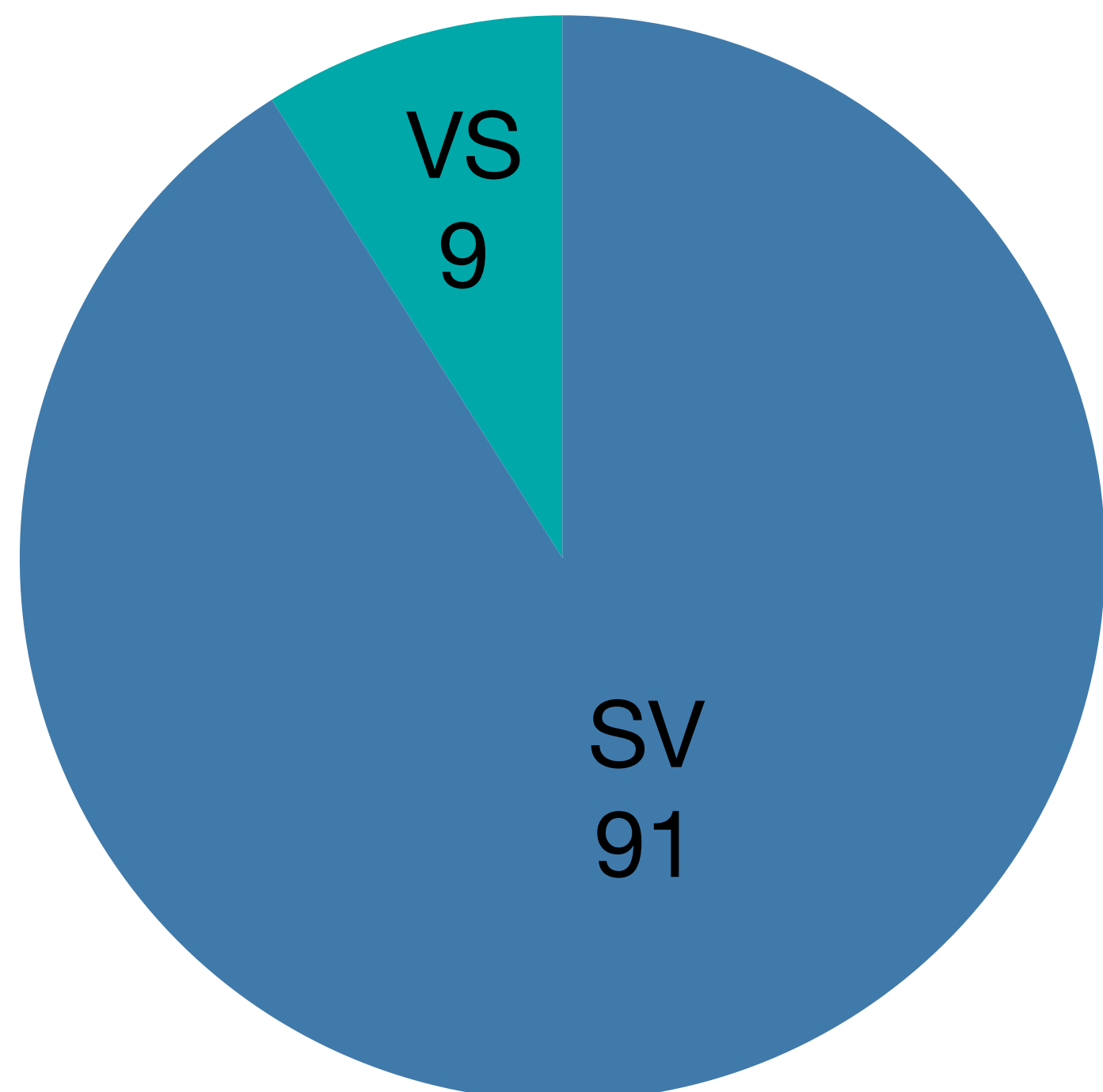
##	<i>nantuwu</i>	<i>ka</i>	<i>nguna</i>	<i>walya-ngka //</i>
##	horse	AUX.PRS	lie.PRS	ground-LOC
##	np.an:s	aux	v:pred	np:l
	009			0027

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

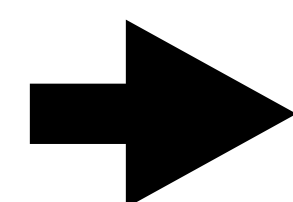
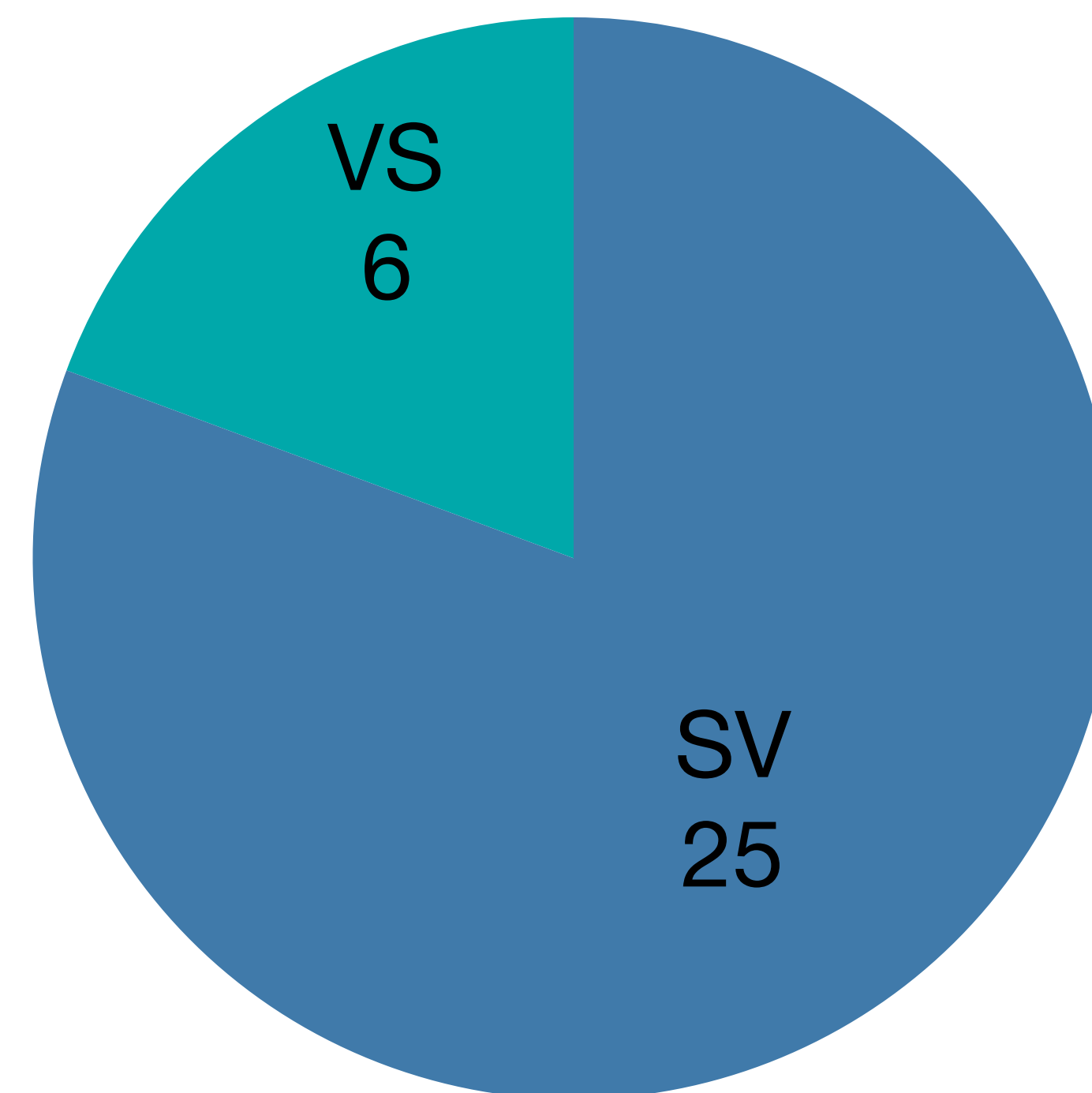
I appreciate any feedback on how I could do more detailed prosodic coding — there will be audio examples in the preliminary results section!

Initial results

Intransitive

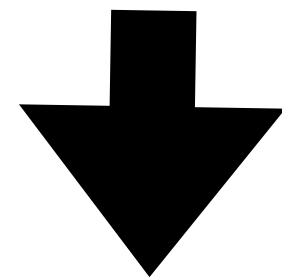


Transitive



What “triggers” VS order?

- (6) *nyina-mi ka malju*
sit-NPST 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

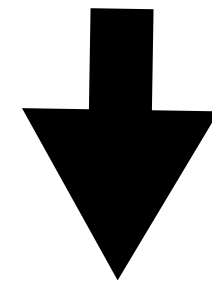
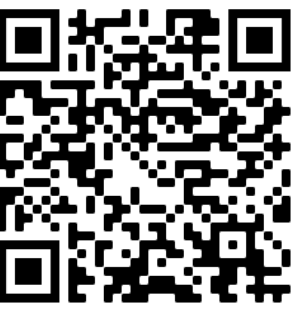
kamina jintan-gku ka // jarntu manu yard-jangka // nyinami ka malju / kurdu witapardu / warluwana / kamina-jarra ka =pala karri-mi / manu jarntu ka nguna-mi warluwana //

one girl gets the dog from the yard. **sits the boy, the little child, by the fire.** the two girls are standing (there) and the dog lies by the fire



Picture Source: O'Shannessy 2004a: 9

(7) *rarralykaji-rla* =*lu* *ya-nu* *wirlinyi* // *wati* *mankurrpa* /
car-LOC AUX.3PL.SBJ go-PST daytrip man a.few
'Went hunting in the car a few men.' (wa32-3-004-005; O'Shannessy)



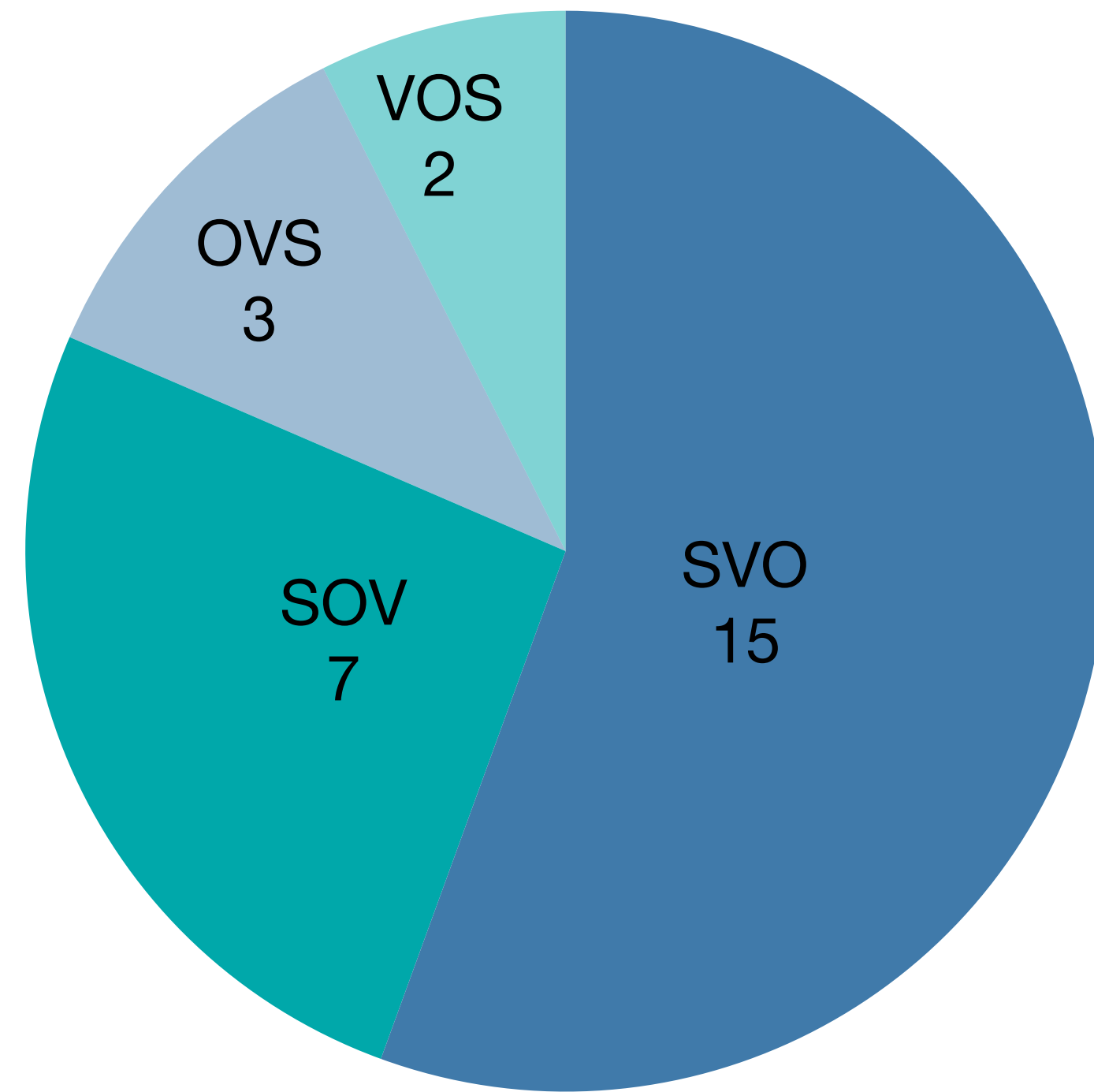
Possible triggers of VS order:

1. Afterthought (= different IU?)
2. Discourse organisation: speaker is unsure?
3. New referents?

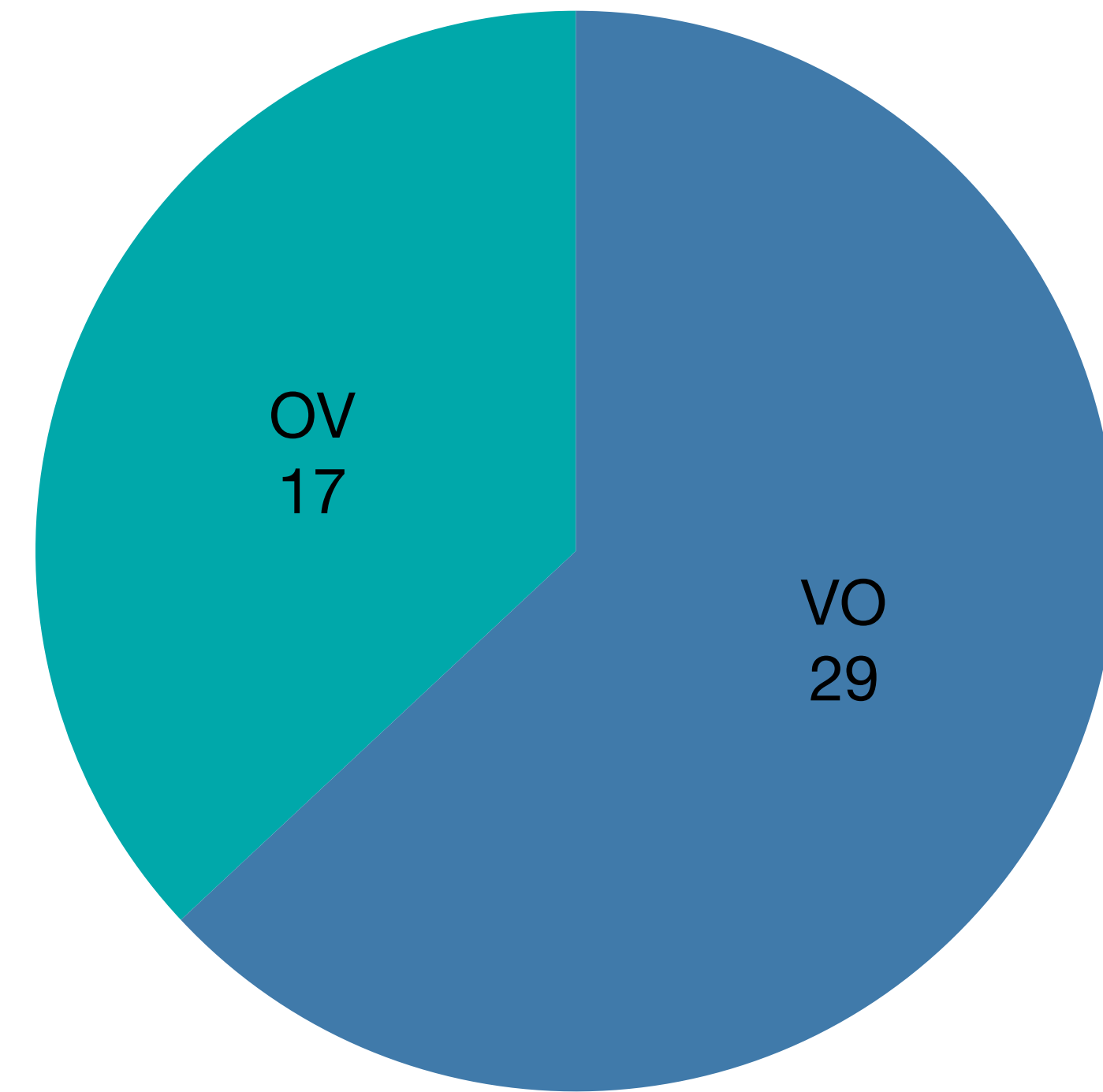


Picture Source: O'Shannessy 2004b: 1

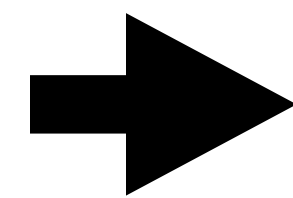
All transitive clauses with overt
subject and overt object



All transitive clauses with overt object



*In finite verbal clauses



What factors play into OV versus VO positioning?

- **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?

Caveat

- (8) *wati-ngki luwa-rnu / puluku*
 man-ERG shoot-PST bullock
 'The man shot the bullock.'
 (wa32-2-008-9; O'Shannessy)



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

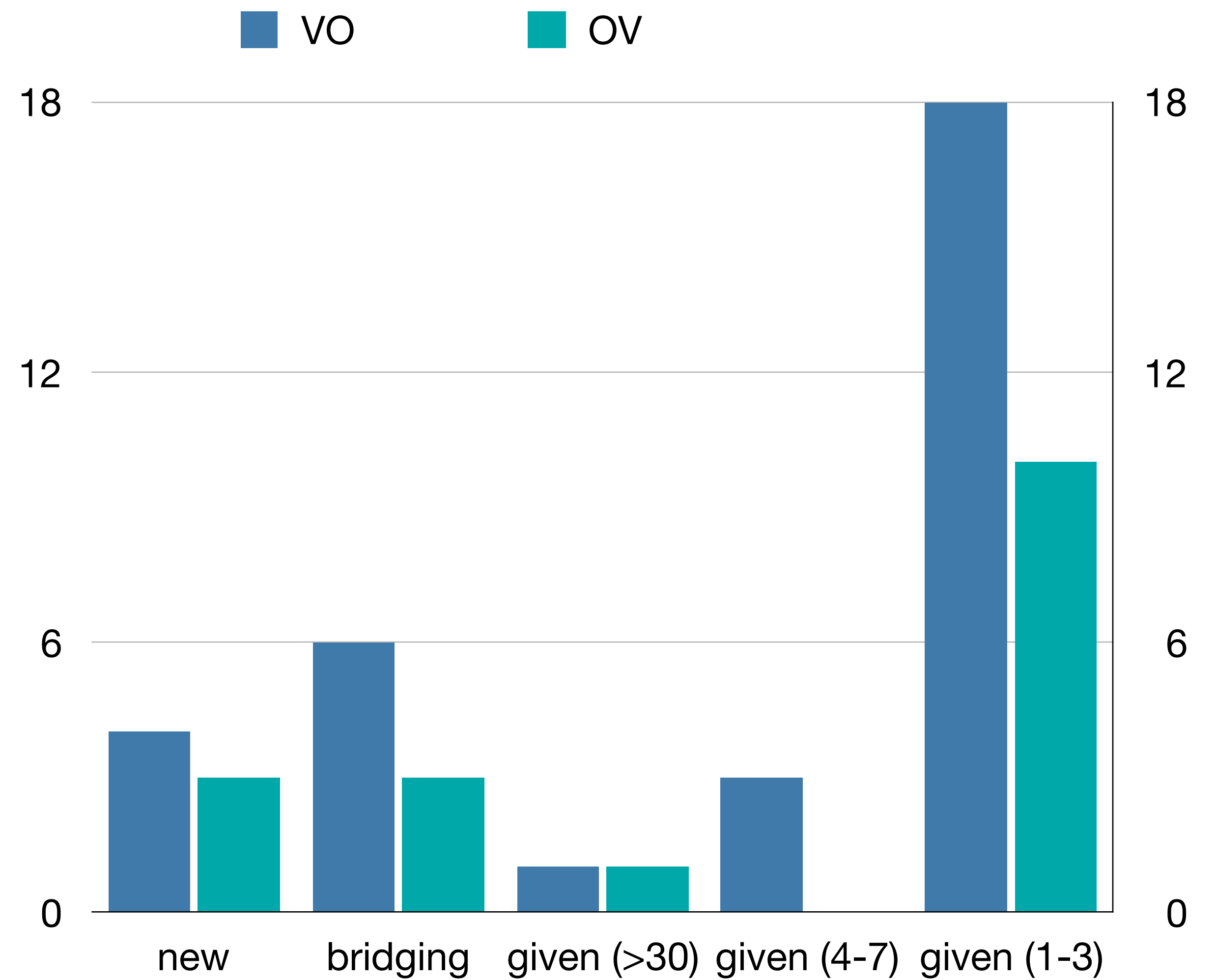
Focus on verb makes VO likely

- Verb before auxiliary in around a third of VO
- Rarely **contrastive** focus, but more focus/prominence/a sequence of events(?)

(9) *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



Given
objects make
VO more
likely

(10) *wirnpa-ngku* *ka* *luwa-rni* *wati*
lightning-ERG AUX.PRS hit-NPST man

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



(11) [...] *nganimpa* *laju* *paji-rni*
we edible.grub pick-NPST


`[...] we pick edible grubs [...]
(j15-034; Nelson 2009)



English loan words make VO more likely

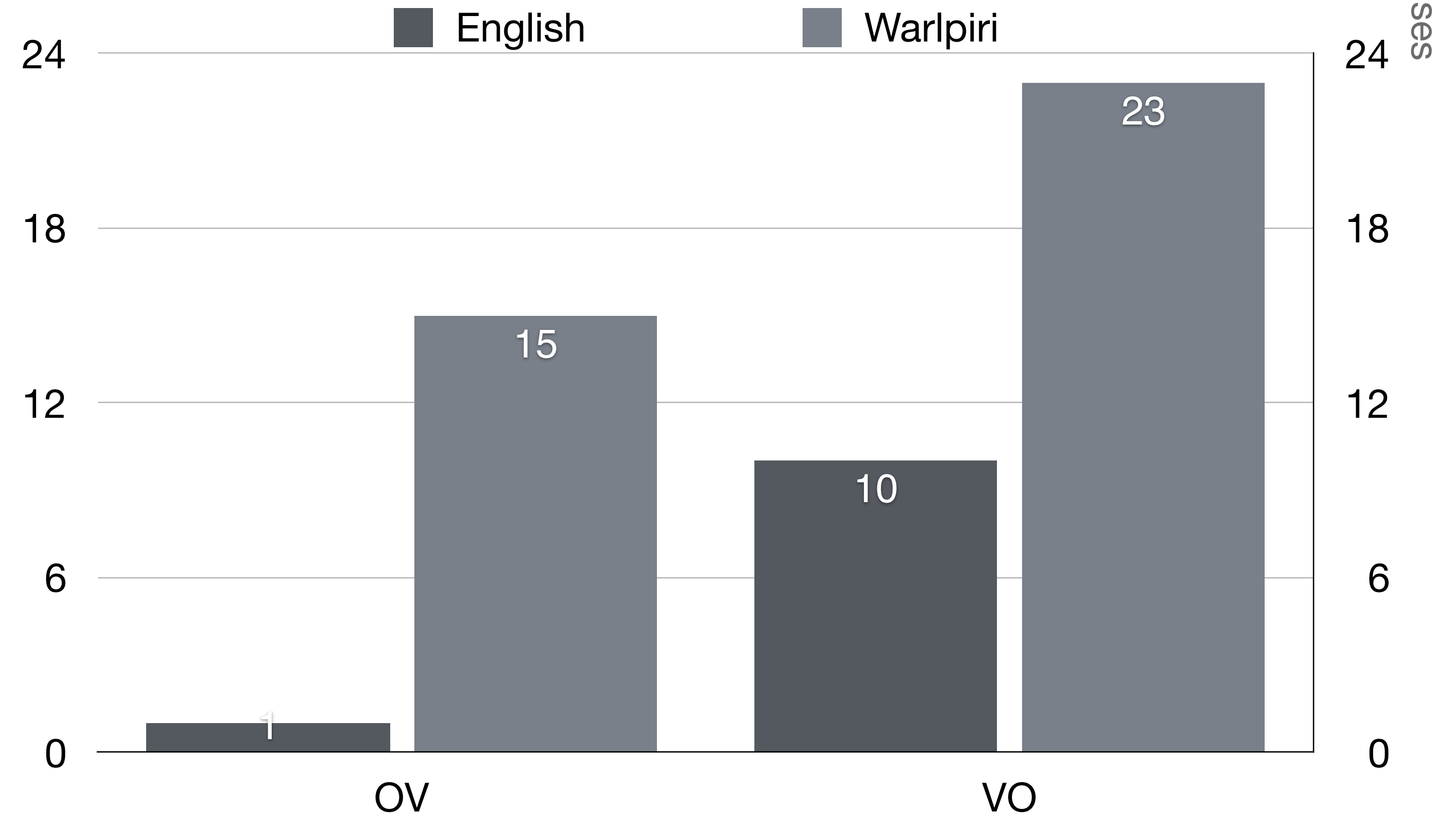
- English loans more frequent in VO (*preliminary*)
- English (SVO) influence?

(12) *wati-ngi manu kurdu-ngu ka*
 man-ERG and child-ERG AUX.PRS
=pala ma-ni fenci
 AUX.3DL.SBJ get-NPST fence



‘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%.

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



- (14) *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

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

What could we discuss?

1. I appreciate any feedback on anything!
2. Am I missing any important factors for word order (that could feasibly be annotated/included in the analysis)?
3. Would it be useful to code **prosody** in more detail, and do you have suggestions for consistent coding?
4. Should I try to make a distinction between VO in a single versus a separate intonation unit? Would it be feasible to assume that an intonation unit break with a specific prosodic contour signals afterthoughts?
5. Might it be worth **coding** semantically weak verbs given my preliminary results?
6. Do you know of any Warlpiri texts collected between 1950 and 1970 in Lajamanu?

References I

Daniels, D. Nampijinpa. 2009. Yakajirri. Warlpiri-patu-kurlangu Jaru (ed.), *Jaru-kurlu. Traditional stories and songs by senior Warlpiri women from Wirliyajarrayi (Willowra), Yurntumu (Yuendumu), Lajamanu and Nyirrpi communities with accompanying audio CD*. Victoria: Trafford Publishing.

Haig, Geoffrey, and Stefan Schnell. 2014. Annotations using GRAID (Grammatical Relations and Animacy in Discourse). Manual Version 7.0. Haig, Geoffrey and Stefan Schnell (eds.), *Multi-CAST: Multilingual corpus of an- notated spoken texts* (multicast.aspra.uni-bamberg.de/) (02. 08. 2019).

Hale, Kenneth. 1983. Warlpiri and the grammar of non-configurational languages. *Natural language and linguistic theory* 1(1), 5-47.

Hale, Kenneth. 1992. Basic word order in two "free word order" languages. Payne, Doris L. (ed.), *Pragmatics of word order flexibility*, Amsterdam/Philadelphia: John Benjamins (= Typological studies in language; 22).

Hale, Kenneth. 2002. Core structures and adjunctions in Warlpiri syntax [5]. *Mouton classics. From syntax to cognition, from phonology to text*. Vol. 1. Berlin, New York: de Gruyter, 117-152.

Hale, Kenneth, and Mary Laughren and Jane Simpson. 1995. 80. Warlpiri. Jacobs, Joachim, Arnim von Stechow, Wolfgang Sternefeld and Theo Vennemann (eds.), *Syntax. Ein internationales Handbuch zeitgenössischer Forschung. / An international handbook of contemporary research*. Halbband 2. Handbücher zur Sprach- und Kommunikations-wissenschaft. Berlin, New York: de Gruyter.

Laughren, Mary. 2002. Syntactic constraints in a 'free word order' language. Amberber, Mengistu, and Peter Collins (eds.), *Language universals and variation*. Westport, CT: Praeger Publishers.

Morton, Lady Napaljarri. 2009a. Track 1 Manja. Warlpiri-patu-kurlangu Jaru (ed.), *Traditional stories and songs by senior Warlpiri women from Wirliyajarrayi (Willowra), Yurntumu (Yuendumu), Lajamanu and Nyirrpi communities with accompanying audio CD*. Victoria: Trafford Publishing.

Morton, Lady Napaljarri. 2009b. Track 2 Janjinngi. Warlpiri-patu-kurlangu Jaru (ed.), *Traditional stories and songs by senior Warlpiri women from Wirliyajarrayi (Willowra), Yurntumu (Yuendumu), Lajamanu and Nyirrpi communities with accompanying audio CD*. Victoria: Trafford Publishing.

Morton, Lady Napaljarri. 2009c. Track 3 Yinpaka. Warlpiri-patu-kurlangu Jaru (ed.), *Traditional stories and songs by senior Warlpiri women from Wirliyajarrayi (Willowra), Yurntumu (Yuendumu), Lajamanu and Nyirrpi communities with accompanying audio CD*. Victoria: Trafford Publishing.

Morton, Lady Napaljarri. 2009d. Track 4 Jitirlpuru. Warlpiri-patu-kurlangu Jaru (ed.), *Traditional stories and songs by senior Warlpiri women from Wirliyajarrayi (Willowra), Yurntumu (Yuendumu), Lajamanu and Nyirrpi communities with accompanying audio CD*. Victoria: Trafford Publishing.

References II

Morton, Lady Napaljarri. 2009e. Track 5 Wapilingki. Warlpiri-patu-kurlangu Jaru (ed.), *Traditional stories and songs by senior Warlpiri women from Wirliyajarrayi (Willowra), Yurntumu (Yuendumu), Lajamanu and Nyirrpri communities with accompanying audio CD*. Victoria: Trafford Publishing.

Nelson, L. Nakamarra. 2009. Track 15 Jujuminyi-minyi. Warlpiri-patu-kurlangu Jaru (ed.), *Jaru-kurlu. Traditional stories and songs by senior Warlpiri women from Wirliyajarrayi (Willowra), Yurntumu (Yuendumu), Lajamanu and Nyirrpri communities with accompanying audio CD*. Victoria: Trafford Publishing.

O'Shannessy, Carmel. 2004a. The monster story. O'Shannessy (ed.), *The monster stories: picture stimuli for elicited production* (unpublished series). Nijmegen: Max Planck Institute for Psycholinguistics.

O'Shannessy, Carmel. 2004b. The bush coconut story. O'Shannessy, Carmel (ed.), *The monster stories: picture stimuli for elicited production* (unpublished series). Nijmegen: Max Planck Institute for Psycholinguistics.

O'Shannessy, Carmel. 2006. *Language contact and children's bilingual acquisition: Learning a mixed language and Warlpiri in northern Australia*. PhD Thesis, University of Sydney, Canberra.

Presley, M. Napurrurla. 2009. Track 7 Yawakiyi. Warlpiri-patu-kurlangu Jaru (ed.), *Traditional stories and songs by senior Warlpiri women from Wirliyajarrayi (Willowra), Yurntumu (Yuendumu), Lajamanu and Nyirrpri communities with accompanying audio CD*. Victoria: Trafford Publishing.

Schiborr, Nils N. & Schnell, Stefan & Thiele, Hanna. 2018. *RefIND — Referent Indexing in Natural-language Discourse: Annotation guidelines*. Version 1.1. (multicast.aspra.uni-bamberg.de/)

Simpson, Jane. 2007. Expressing pragmatic constraints on word order in Warlpiri. Zaenen, Annie, Jane Simpson, Tracy Holloway King, Jane Grimshaw, and Joan Maling (eds.), *Architectures, rules, and preferences; variations on themes by Joan W. Bresnan*. Stanford, California: CSLI Publications.

Simpson, Jane, and Ilana Mushin. 2008. Clause-initial position in four Australian languages. Mushin, Ilana, and Brett Baker (eds.), *Discourse and grammar in Australian languages*, 25-57. (= Studies in Language Companion Series; 104)s

Simpson, Jane. 2012. Information structure, variation and the Referential Hierarchy. F. Seifart, G. Haig, N. P. Himmelmann, D. Jung, A. Margetts and P. Trilsbeek (ed.), *Potentials of Language Documentation: Methods, Analyses, and Utilization*, University of Hawaii Press, Honolulu, 73-82.

Swartz, Stephen. 1991. Constraints on zero anaphora and word order in Walrpiri narrative text. *SIL - AAIB Occasional Papers No.1*. Australian Aborigines and Islanders Branch, Summer Institute of Linguistics, Darwin.

Vollmer, Maria. to appear. Multi-CAST Warlpiri annotation notes. Haig, Geoffrey and Stefan Schnell (eds.), *Multi- CAST: Multilingual corpus of annotated spoken texts* (multicast.aspra.uni-bamberg.de/)).

Acknowledgements

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. Jane has put in a lot of time and effort to give me feedback on these slides and my PhD project in general, and I thank her for her patience and deep insight. I also want to thank Uta Reinöhl for her support for and valuable feedback on my PhD thesis.

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)
maria.vollmer@linguistik.uni-freiburg.de
 @vollmer_maria