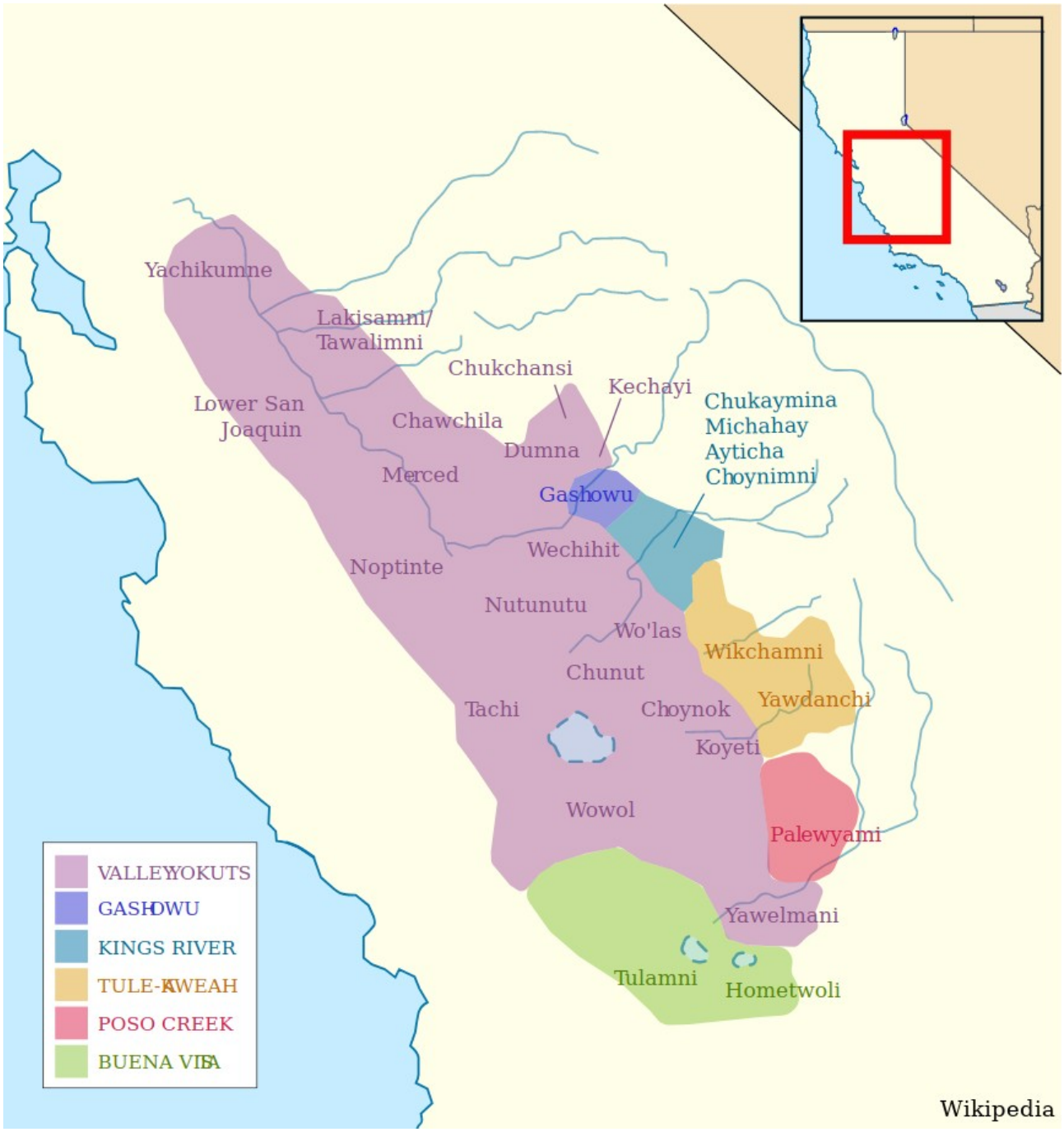
1/12 @wieldorg #FragmentedLanguageWorkshop2020
Yowlumne (aka Yawelmani) is a Yokuts language of the Valley Yokuts branch native to California. It appears near the bottom right of the large purple area of the map.



2/12 Yowlumne has a complicated system of secundative alignment in ditransitive constructions (Weigel 2005). Secundative alignment is where the recipient-like argument (R) of a ditransitive verb patterns like the patient-like argument (P) of a transitive verb (MHC 2010).

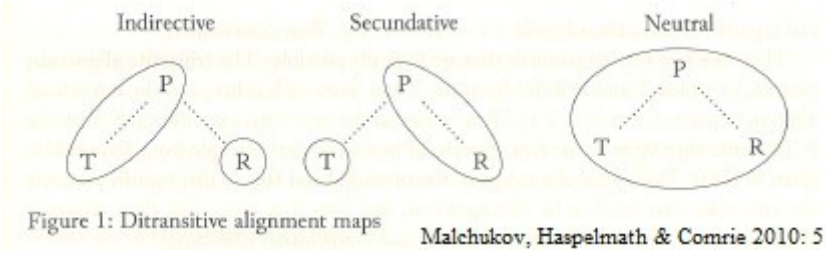
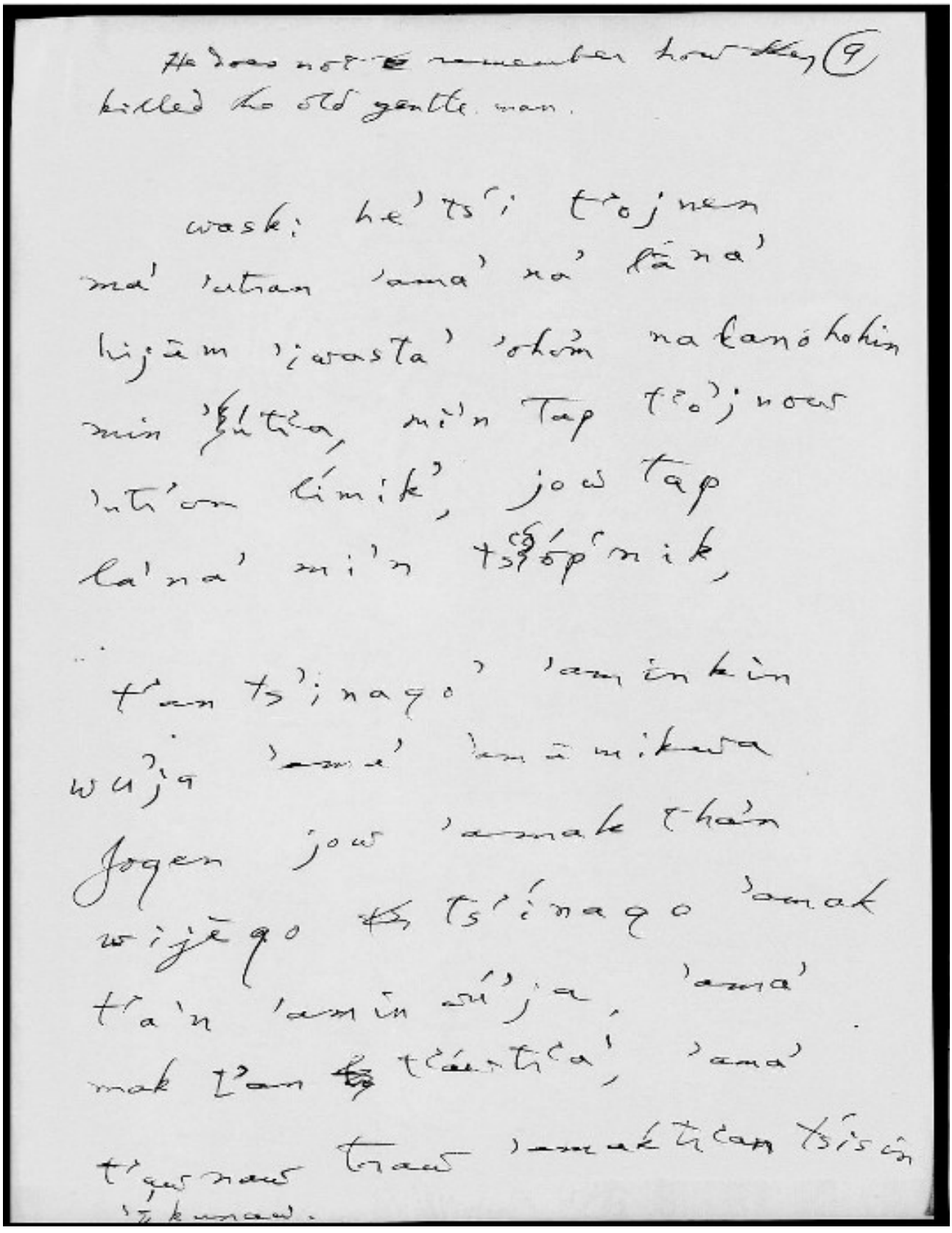


Figure 1: Ditransitive alignment maps Malchukov, Haspelmath & Comrie 2010: 5

3/12 Large amounts of available data that document the alignment system in Yowlumne are found in messy, handwritten archival documents in the archives of John Peabody Harrington and Stanley Newman, recorded between 1910 and 1940. This data is not readily searchable in this form.



4/12 Digitization and accessibility to these documents takes place in several stages. The first is digital transcription of 17 texts (so far).

```
Slides 1145-1154 has a story (translation is slide 1154-1163)

mi'n 'aman hutr'on. wija'n
pajās hilḗtits',
limi 'aman tr'owan.
'ama' tr`at lakli
t`awin mitran jow kíjem,
limik'in tr'awin. 'ama'
jet' jow thawinmi laklēn
jow kijhin, 'ama'
hewanuk má'an waki
man tr'áwán'an,
wáski wija'n k`a'jiw:
mi'n mantran k`ojwen,
mi'natran ti'sen wi'ja',

'ama' k`a'jiw
t`anintr`aw hija'm, mokjowmen
limk'in. t`ahak nim tran
toḵojni, k`ajutrina'n,
limk'in mokji, k`aju tap
tisisan limik' [changed himself to lim.], mi'n tran
jṓman k`a'jiw, jow traw
'ohóm. qoqó'traw, wija'n
k`a'jiw, jow traw qoqomakin
'ótqo, hulom 'ohjuk,
wija'n tr`an, 'ama' tr`aw
thanhin tr`aw t`auwtran
maqitswij tóḵój nitr`an,
'ama' t`awtran hetam makmithaw
tawtran wōjon,
```

5/12 The second stage is a preliminary text normalization process using an algorithm involving Levenshtein distance (image 1). Harrington confused many sounds in his transcriptions, so I use sound classes (image 2) to group these sounds in the Levenshtein distance calculations.

$$\operatorname{lev}(a,b) = \begin{cases} |a| & \text{if } |b| = 0, \\ |b| & \text{if } |a| = 0, \\ \operatorname{lev}(\operatorname{tail}(a), \operatorname{tail}(b)) & \text{if } a[0] = b[0] \\ 1 + \min \begin{cases} \operatorname{lev}(\operatorname{tail}(a), b) \\ \operatorname{lev}(a, \operatorname{tail}(b)) \\ \operatorname{lev}(\operatorname{tail}(a), \operatorname{tail}(b)) \end{cases} & \text{otherwise.} \end{cases}$$

Wikipedia

k k' g kˋ kh k̲ k̄ ḳ k’ kː
t t' d tˋ th t̄ t̲ ṭ
r tr tr' dr trˋ tʃˋ ṭ tʃ ṭ’ ṭ ṭʿ t̲r̲ tʃrˋ tʃr
c ts tsˋ dz zh chr ts' c' ʦ̣ tc tc’ ʒ ç ʦ tʂ tʂˋ t̄ʂ tʂ
p p' b pˋ ph p̄ pʿ p_
m m̄ ḿ mː
n ñ n̰ n̲ ṇ
x q q̄ g q̇ q̣
l l̲ ɬ
s š ʃ shr s s̲ sh ʂ sː ʂ
h h̰
a áː á ā á a̰ ă ɑ ą aː á ạ a̲ ā˘ ɑ ə à a🅰 ā aˋ
e é ē ɛ ẹ eː ḗ ɛ́ éː ḛ ɛ́ e̲ éː ɛː éː ɛ̲ ĕ eˋ
i í ī ǐ iː íː ǐ ı̨ i🅰 iˋ
o ọː ó ō ǒ ǒ ọ oː ɔ́ ɔ ɔ́ː ɔɔ ɔː ɔ̨ ɔ́ ʌ ʌ́ ɔ̲ ọ ʊ́ ʊ ò ȯ oˋ
u ú ū ụ uː uʿ ụ ű ŭ u🅰 ű ū uˋ
w w u̲ w̲
y j íʿ i̲ yː j̄
' ' ‘ ’ ': ‘ː ’ː '🅰

6/12 This first pass at normalization is then hand-checked and corrected for accuracy, with a percentage given to indicate subjective certainty for the word identified.

```
mi'n    mi'in   'soon'  100.0%
'aman   'aman   '3PL.NOM'       100.0%
hutr'on hotr'on 'hotron game'   100.0%
.       .       [punc]  100.0%
wija'n  wiyi    'say, do'       100.0%
pajās   [unknown]       '[unknown]'     0.0%
hilētits'       hile:c'ic'      'hilēcic'       100.0%
,       ,       [punc]  100.0%
limi    limik'  'prairie falcon'        100.0%
'aman   'aman   '3PL.NOM'       100.0%
tr'owan tr'aw'a 'win from'      100.0%
.       .       [punc]  100.0%
'ama'   'ama'   'and'   100.0%
tr`at   tra:nit 'that.SG.ABL'   25.0%
lakli   lakli'  'different'     100.0%
t`awin
mi%     tawin   'to become morning'     100.0%
tran    tran    'that.SG.PRIM'  100.0%
jow     yow     'and, also, again'      100.0%
kíjem   gi'iy   'oppose'        50.0%
,       ,       [punc]  100.0%
limik'in        limik'  'prairie falcon'        100.0%
tr'awin tr'aw'a 'win from'      100.0%
.       .       [punc]  100.0%
```

7/12 This first-pass normalized data is used to train a Transformer model (Vaswani et al. 2017) that maps from characters to lexemes. The Transformer model can then be used to normalize other data obtained from the archives.

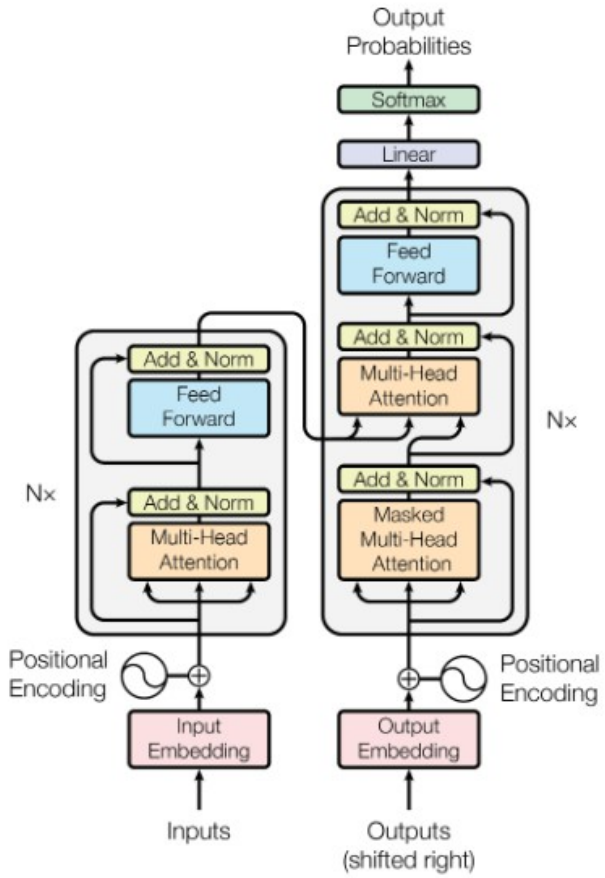
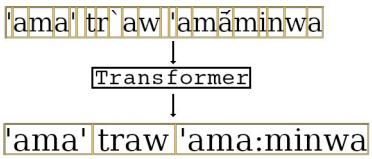
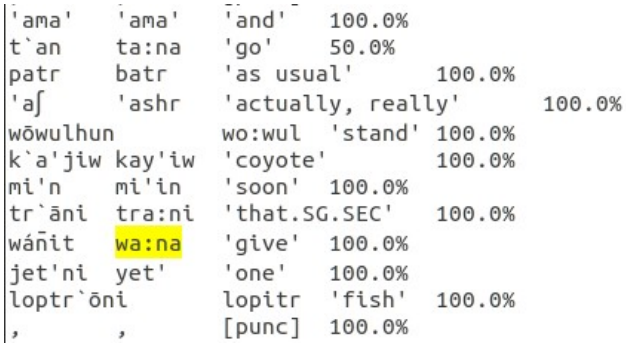


Figure 1: The Transformer - model architecture.

Vaswani et al. 2017: 3



8/12 The data normalized by the Levenshtein distance calculations and the Transformer model can be searched by lexeme. This enables selecting individual ditransitive verbs (e.g. *wa:na ‘give’) to chart out their attested argument structures, including in passive environments.



'ama'	'ama'	'and'	100.0%	
t`an	ta:na	'go'	50.0%	
patr	batr	'as usual'	100.0%	
'aʃ	'ashr	'actually, really'	100.0%	
wōwulhun	wo:wul	'stand'	100.0%	
k`a'jiw	kay'iw	'coyote'	100.0%	
mi'n	mi'in	'soon'	100.0%	
tr`āni	tra:ni	'that.SG.SEC'	100.0%	
wáñit	wa:na	'give'	100.0%	
jet'ni	yet'	'one'	100.0%	
loptr`ōni	lopitr	'fish'	100.0%	
,	,	[punc]	100.0%	

9/12 The data obtained provide preliminary results for ditransitive alignment with passivization: the examples show *wa:na ‘give’. The primative-marked argument in the active sentence is the recipient, while the secundative-marked argument is the theme in active and passive.

mi'n	'amānwa	sōk'onni	wánin
mi’in	‘ama:nwa	so:konni	wanhin
mi’in	‘ama:nwa	so:kon-ni	wa:n-hin
soon	3PL.PRIM	tobacco-SEC	give-AOR
‘soon he gave them tobacco’ (text 10, 132-135)			

mi'n	tr`āni	wánīt	jet'ni	loptr`ōni
mi’in	ṭʰa:ni	wa:nitʰ	yet’ni	lopʰṭʰo:ni
mi’in	ṭʰa:ni	wa:n-itʰ	yet’-ni	lopʰṭʰ-a:ni
soon	that.SEC	give-PAOR	one-SEC	fish-SEC
‘soon he (=Coyote) was given that one fish’ (text 10, 87-91)				

10/12 Applicativized verbs work the same way, as with *k’o’o ‘throw’. For the base verb, the theme (T) is primative-marked. The applicative demotes T to secundative, and the recipient-like argument (R) is primative. In the passive, R is subject, and T remains secundative.

k'ohin	tran	'oʃto	'éntr`am
k’o’hin	ṭʰan	‘oṣtʰow	‘enṭʰam
k’o’-hin	ṭʰan	‘oṣtʰ-aw	‘enṭʰam
throw-AOR	that.PRIM	fire-LOC	sleeping.potion
‘he threw that sleeping potion in the fire’ (text 5, 287-290)			

'ama'	'amintr`an		ts'imēk'āt`aw	k'o'sithin	tr`āni	wiʃats'ni
‘ama’	‘amin	ṭʰan	c’ime:k’a:tʰaw	k’o’sitʰhin	ṭʰa:ni	wiṣac’ni
‘ama’	‘amin	ṭʰan	c’imik’-ᴱᶜa:-tʰaw	k’o’-sitʰ-hin	ṭʰa:ni	wiṣa:c’-ni
and	3SG.GEN	that.PRIM	close.eyes-INCH-NDG	throw-APPL-AOR	that.SEC	arrow.straightener-SEC
‘and when he_i closed his_i eyes, he_j threw the arrow straightener at him_i’ (lit. ‘and at his_i closing of eyes, he_j threw-at him_i the arrow straightener’) (text 6, 400-406)						

'amamak'o'sitnit			tr`ani	'aminōkun	ʃunāni
‘ama’	ma’	k’o’sitʰnitʰ	ṭʰa:ni	‘amino:gun	ṣuna:ni
‘ama’	ma’	k’o’-sitʰ-nitʰ	ṭʰa:ni	‘amino:gun	ṣun-Ø-a:ni
and	2SG.NOM	throw-APPL-PFUT	that.SEC	3PL.GEN	stuff-VN-SEC
‘and you will be thrown-at that stuffing(?) of theirs’ (text 9, 119-124)					

11/12 In either case, the preliminary results show that passivization applies to the recipient-like argument (R): R becomes subject. The theme remains a secundative-marked argument regardless. This is true for both ditransitive verbs and applicative-marked verbs.

*wa:na ‘give’

Active

(Agent)	Recipient-like	Theme
(NOMINATIVE)	PRIMATIVE	SECUNDATIVE

Passive

(Recipient-like)	Theme
(NOMINATIVE)	SECUNDATIVE

*k’o’o ‘throw’ + applicative -sit^{h}-

Active

(Agent)	Recipient-like	Theme
(NOMINATIVE)	PRIMATIVE	SECUNDATIVE

Passive

(Recipient-like)	Theme
(NOMINATIVE)	SECUNDATIVE

12/12 In conclusion, the computational approach used to access the data will be useful to those working with archival data. This approach enabled analysis of the ditransitive alignment in Yowlumne and found that passivization promotes the primative-marked argument to subject.

References

Malchukov, Andrej, Martin Haspelmath & Bernard Comrie (MHC). 2010. *Studies in ditransitive constructions: A comparative handbook*. Berlin: De Gruyter Mouton.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser & Illia Polosukhin. 2017. Attention is all you need. Manuscript presented at the 31st Conference on Neural Information Processing Systems.

Wiegel, William Frederic. 2005. Yowlumne in the Twentieth Century. UC Berkeley PhD dissertation.