



HAL
open science

Silent gaps as a distinctive feature of direct speech: a corpus study of Nisvai narratives, Vanuatu

Jocelyn Aznar

► **To cite this version:**

Jocelyn Aznar. Silent gaps as a distinctive feature of direct speech: a corpus study of Nisvai narratives, Vanuatu. The grammatical variation of reported speech across languages: A data-oriented workshop, Nov 2021, Helsinki, Finland. hal-03457980

HAL Id: hal-03457980

<https://hal.science/hal-03457980>

Submitted on 30 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Leibniz-Zentrum
Allgemeine Sprachwissenschaft

SILENT GAPS AS A DISTINCTIVE FEATURE OF DIRECT SPEECH

A CORPUS STUDY OF NISVAI NARRATIVES, VANUATU

Jocelyn Aznar

February 11, 2021

Leibniz ZAS

TABLE OF CONTENTS

1. Introduction

About Nisvai

Corpus

Methodology

2. Reported Speech in Nisvai

Indirect Speech

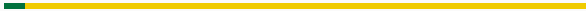
Direct Speech

3. The silent gaps of Nisvai Direct Speech

Silent gaps

4. Conclusion

INTRO



Study of silent gaps as part of the Nisvai Direct Speech device

- how silent gaps contribute to the Nisvai's Direct Speech device,
- developing tools for doing cross-linguistic comparison of silent gaps.

THE NISVAI COMMUNITY AND CORPUS' PROJECT

An oral language which has been the subject of only one previous lexical study by Charpentier (1984).

Around 200 natives speakers: the native language of the children.

The language is used in most daily activities, except:

- Relationships with other village communities: Bislama.
- Church: Bislama, a bit of English. Translating to Nisvai.
- School: Teaching language is French, but children actually learn to speak Bislama fluently there, and just few words of French.

Corpus original purposes:

- Ethnolinguistics research on Nisvai narrative practices,
- For the Nisvai speakers: having a written system and language resources for the local school.

- Austronesian language>Oceanic Language> Central Vanuatu¹
- SVO language
- Offers some lexical flexibility (van Lier, 2016), but not omnipredicative² as other languages of the area.
- Very few adjectives and adverbs, but many stative verbs (including numbers, color terms):
- Only animate objects can be subject and no passive.
- Tense: non Future/ Future
- Mood: Irrealis/Realis

¹See <https://glottolog.org/resource/languoid/id/nisv1234> for a more detailed description.

²Many Austronesian language are described as omnipredicative (Launey, 2004) (François, 2001).

³The name, used by the speakers, means "what", as it is very common in the area.

The corpus of Nisvai narratives:

- 165.5 minutes with 32 narratives of between 0.67 to 13.32 minutes,
- 3041 annotations units : 1301w, 1740m,
- 18 speakers, 2 sex (5w, 13m), from 5 to 80 years old,
- practical orthography: no characters outside the ASCII set,
- Nisvai speakers regard their narratives as true stories.

ANNOTATING THE NARRATIVES

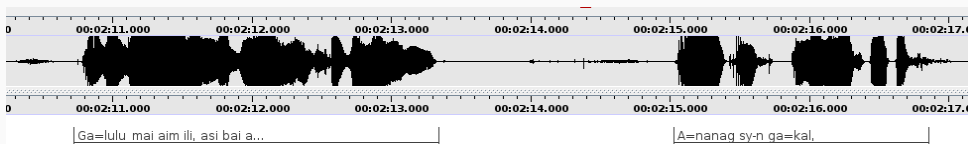


Figure 1: Segmentation into annotation units: interpausal units.

ANNOTATING THE NARRATIVES

00:05:57.250

Mode de lecture

Intervalle de temps: 00:05:55.640 - 00:05:57.250 1610

00:05:56.000 00:05:57.000

Volume 100

Taux 100

Paramètres

Lecteur automatique du média

Créer les annotations manquantes

Montrer le nom des acteurs

Colorier les numéros de colonnes uniques

Parcourir la colonne

Centrer l'affichage temporel sur l'annotation

Configurer...

No	Type 1 : Transcription	Type 2 : Textualité	Type 3 : Référence	Type 4 : Morphologie	Type 5 : Note	Type 6 : Traduction
93	Ga=kai verah van :	pénpétie_9	T21.1201.3.93	35G=dre dernière ASP.A		É dit, dernière :
94	<daa, nangac=lu nahre moq, ga=lu lu mal, daru.	pénpétie_9	T21.1201.3.94	INTER.NEG 15G=porter=rentrer enfant ASP.F 25G=rentrer-RED revenir 1INCL	parcudi : karem bak	<Mais non, [ai déjà ramené les enfants, reviens, allons-y.
95	Be naly:n ga=sab mi cab.	pénpétie_9	T21.1201.3.95	COO.OP entrailles 25G 35G=être_en_colèreaussi INT		Ne sois pas en colère.
96	Ga=me daru, dara=lu van dry va=roh, dara=lap lapus ga=uu moq bul...	pénpétie_9	T21.1201.3.96	25G=venir 1INCL 1INCL=rentrer aller COO aller=restér 1INCL=RED=jouer 35G=être_bien ASP.F COO.OP		Viens, allons-y, revenons et restons jouer ensemble, c'est bien comme ça mas...
97	ga=pac=lu a=nahre moq dry van va=lig+1, Daru dara=van va=vhasus=hni. >	pénpétie_9	T21.1201.3.97	25G=apporter=rentrer ART.P=enfant ASP.F COO aller aller=déposer-RTR 1INCL 1 INCL=aller aller=nourrir=OB]	vhasus hni : donner le sein.	raporte les enfants et va les déposer. Allons les nourrir. >
98	Bil a=Lmwume, nanovin sy:n ga=sab.	pénpétie_10	T21.1201.3.98	COO.OP ART.P=NON penser-NOM PA.35G 35G=mauvais		Mais le pigeon est en colère.
99	Ga=kai ga=mas (ga=roh dry) mac m.	pénpétie_10	T21.1201.3.99	35G=dre 35G=devoir 35G=restér COO mourir INT		É se dit qu'il dot mourir.
100	Ga=cahn mías, ga=dryb ay:n+1 Barbaru.	pénpétie_10	T21.1201.3.100	35G=jeter-INTR à nouveau 35G=tomber INDEF-INTR LIEU	barbaru : blaksand / nabangura.	É en jette une à nouveau, elle tombe dans un endroit comme Barbaru.
101	Ga=kai dry svoh.	pénpétie_10	T21.1201.3.101	35G=chanter COO voler		É chante et vole :
102	<sa-m natu-m aru meh-mah, i-sa-q natu-q aru vog-vog>	chant_5	T21.1201.3.102	PRO-PA-25G enfant-25G 3PLASP.EC être_noir-RED PRO-PA-15G enfant-15G 3PLASP.EC être_marron-RED		<Les noirs, tes deux enfants sont noirs, les noirs, mes deux enfants, sont marrons.>
103	Ga=svoh dry...	pénpétie_10	T21.1201.3.103	35G=s'envole COO		É s'envole et
104	ga=kai mías van van bulaha qanag	pénpétie_10	T21.1201.3.104	35G=chanter à nouveau ASP.A.MUL attendre comme		chante à nouveau, l'attend comme ça.
105	pac naroh ay:n m, cah ni ga=dryb libagar,	pénpétie_10	T21.1201.3.105	prendre chose DEF à nouveau jeter-INTR 35G=tomber plage		prend une autre liane, la jette et elle tombe sur la plage.

Figure 2: Content annotation using ELAN

Computer notebook: Jupyter (Kluyver et al., 2016))

- Reproducibility Berez-Kroeker et al. (2018),
- Facilities for doing exploration studies using quantitative and visualisations analysis,
- Algorithmic studies: apply queries on each annotation and have a result,
- Helps for replication as well by providing the algorithms that were used Drummond (2009).

The notebook and the data associated with this presentation will be available soon.

REPORTED SPEECH IN NISVAI

GENERAL REMARKS

Direct speech sequence/frame (Genetti, 2014) = introductory segment + reported speech + conclusive segment

Indirect speech sequence = introductory segment + reported Speech

Indirect Speech	Direct Speech
Keep the narrative's language	Language of the reported speech
Current situation's mood	Mood of the Reported situation
Current situation's deictics	Reported situation's deictics
No Enactment	Enactment ⁴

⁴different voice timbre, different pitch

Indirect speech is often used to convey the thoughts of the characters.

(1) Interior speech, T21.t2013.98-99

Bul a=Livwume, nanov-in sy-n ga=sab.
COO.OP ART.P=NOM think-NOMI PA-3SG 3SG=bad

‘But the Pigeon is angry.’

Ga=kal ga=mas mac ni.
3SG=say 3SG=must die INT

‘She is telling herself that she has to die⁵.’

⁵Little enactment here: lower and angry voice

(2) Indirect speech, T15.2013.10

Ara=lav-lavis van ili, bog a=ckai, tataq
3PL=RED-play ASP.A ASP.R day 3SG=be_one father

sy-n ga=pac-i ga=kal ga=va=vor lhoi.
PA-3SG 3SG=take-INTR 3SG=say 3SG=go=draw_water river

'They play sometimes, one day, his father carries something
and says he is going to draw water in the river.'

(3) Interior speech, T1.2013.125

Bul kabani nahemac ili ari avy n
COO.OP group deamon DET 3PL DEF 3PL=be
ara=tuv lyn bar
INST mud DET

ili

PRO 3SG=say 3SG-?? hurt-INTR PRO 3SG=say 3SG-?? hurt-INTR
avy n ga=kal ga=emo tams-i avyn ga=kal ga=emo tams-i.

'But among the group of the deamon that went to the mud,
one says he will hit him, another says he will hit him.'

435 direct speech occurrences in the corpus.

Introductory verbs:

- kal: "to say", not mandatory but... almost.
- tog-i: "to listen-to something",
- qan: "to be like",
- kai : "to sing"

Interjections or Appellative: Marking the beginning of the quotation or turn-taking in a dialogue.

Poetic sequences: The poetic sequences are always introduced as direct speech.

(4) "pus": support verb for direct speech, **T44.2015.37**

ara=pus ga=kal: «Haig, nanag sa-m nmah a?»
3PL=ask 3SG=say 2SG mother PA-2SG snake INTER

'They ask, s/he say: «You, your mum is a snake, eh!»'

(5) T21.t2013.43

«Bul ruben isy-n ga=meh-meh avyn
 COO.OP PREP.CAUSE PA-3SG 3SG=to_be_black-RED REL
ga=meh-meh.» Bul gu=r=lys sa
 3SG=to_be_black-RED COO.OP 3SG.ASP.EC-FUT=see PA
Livwume, dry tog-i: «Ruben kal isy-n
 BIRD COO think-INTR PREP.CAUSE say PA-3SG
redgu-n ursao ga=vog-vog ag
 throat-3SG DEI.LOC.E 3SG=to_be_brown-RED DEI.P
qan, isy-n ga=vog-vog qan
 to_be_like PA-3SG 3SG=to_be_brown-RED to_be_like
ag.» dry tog-i
 DEI.P COO.VB think-INTR

‘«But hers are black, very black.» And he looks at those of the Pigeon and thinks: «Why she says hers have a brown throat like that, hers have a brown throats like that.» he thinks.’

(6) T41.2015.97 T41.2015.101

«E!», *ga=kal-i visgy-n napnevur sy-n aru*
 INTER 3SG=say-INTR OBJ-3SG woman PA-3SG 3PL.DU
ga=kal : «E!
 3SG=say INTER

‘«Eh» he says to his two wives: «Eh !’

Mura=ls-i, avys ara=mac-mac pal moq, amru
 2PL=see-INTR other 3PL=be_dead ASP.T ASP.F 2PL
ni mycig mra=roh-roh.
 INT COO.TMP 2PL=stay-RED

‘You see, the others are dead already, only you two are left.’

Be mra=kaikai nahsy-q,[...]
 NEG 2PL=call name-1SG

‘Don’t call my name.[...]’

(7) Silent gap just after the interjection, **T1.2013.189-190**

Bul *a=Tamata* *van* *va=puc* *ari* *kal:* «Eh,
COO.OP ART.P=NOM go go=lie 3PL say INTER

‘So Tamata goes to lie and says: «Eh,’

dara=qamqam *ni.»»*

1INCL=to_speed_up INT

‘let’s go fast.»’

They represent 112 of 435 direct speech occurrences in the corpus.

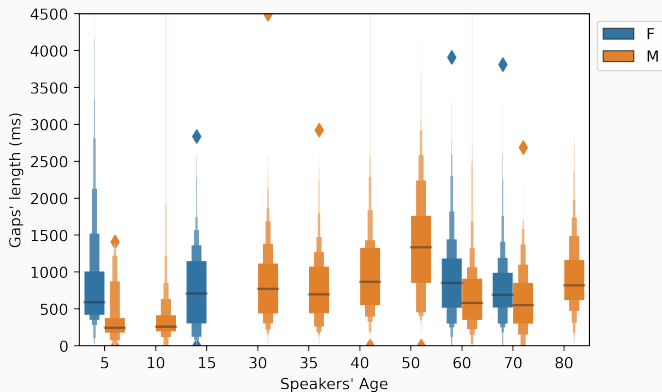
THE SILENT GAPS OF NISVAI DIRECT SPEECH



GENERAL REMARKS

Breath gaps included into silent gaps.

	25%	mean	50%	75%
% of silence ⁶	20.6	25,9	26,6	30
length	430	845	745	1100



What are the factors that influence the length of silent gaps:

- Sex (coef +- 59.5137, $P > |t| = 0,000$)
- Age (coef 1.7553, $P > |t| = 0.003$)
- Position in the text (coef -2.5702 ; $P > |t| = 0.000$)
- Local Speech Rate⁷ (coef 15.0498 ; $P > |t| = 0.000$)
- Text's length (coef 0.3552 ; $P > |t| = 0.000$)

According to the statistical tests (OLS), these variables affects significantly the length of silent gaps, but only very slightly (as the coefficients show.).

⁷Local Speech Rate is calculated by counting the number of phonemes in an annotation units.

A MODEL OF NISVAI'S DIRECT SPEECH SEQUENCE

Direct Speech	Verb	gap	INTER	gap	RS	gap
Frequencies	? ⁸	199	356 ⁹	112	435	417

⁸I have not counted them yet, but direct reported speech without introductory verbs exists, but are not numerous.

⁹Only 14 interjections are not used for marking the beginning of direct speech. No use of interjection for indirect speech.

SILENTE GAPS AND DIRECT SPEECH: BEFORE AND DURING.

Sample: 435 of Direct Speech sequences over around 3000 silent gaps.

The length of the silent gaps :

- after an interjection (112 occurrences): no significant variation ($P > |t|$ 0.545),
- before direct discourse (199 occurrences): no significant variation ($P > |t|$ 0.100),
- after direct reported speech (417 occurrences): significant increase ($P > |t|$ 0.000), mean length 985 ms (compared to 821 ms).

CONCLUSION

in Nisvai, interjections and silent gaps (before and during) are distinguishing Direct Speech from Indirect Speech.

Silent gaps can be after the introductory verbs, but also after the interjection starting the direct speech.

- Nisvai has a very limited number of introductory verbs
- Direct speech is produced using two terms (usually verbs):
 - A verb or a term of manner,
 - A verb introducing the content of the reported speech.

- Taking into account the role of appellation for direct speech.
- Using this as a framework for cross corpora comparison.
- sex and age will not be comparable using language documentation corpora as they mostly contain texts performed by few speakers.

REFERENCES

Andrea L Berez-Kroeker, Lauren Gawne, Susan Smythe Kung, Barbara F Kelly, Tyler Heston, Gary Holton, Peter Pulsifer, David I Beaver, Shobhana Chelliah, Stanley Dubinsky, P Meier, Nick Thieberger, Keren Rice, and Anthony C Woodbury. Reproducible research in linguistics: A position statement on data citation and attribution in our field. *De Gruyter*, 56(1):18, 2018. doi: 10.1515/ling-2017-0032.

Jean-Michel Charpentier. *Atlas Linguistique Du Sud-Malakula (Vanuatu)*. Coll. Langues et Cultures Du Pacifique, 2. SELAF, Paris, 1984.

- Chris Drummond. Replicability is not Reproducibility: Nor is it Good Science. *Proc. of the Evaluation Methods for Machine Learning Workshop at the 26 th ICML*, page 4, 2009.
- Alexandre François. *Contraintes de Structures et Liberté Dans l'organisation Du Discours. Une Description Du Mwotlap, Langue Océanienne Du Vanuatu*. PhD thesis, Université Paris IV–Sorbonne, 2001.
- Carol Genetti. Direct speech reports and the cline of prosodic integration in Dolakha Newar. *Himalayan Linguistics*, 10(1), August 2014. ISSN 1544-7502. doi: 10/ghfbmz.

- Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Matthias Bussonnier, Jonathan Frederic, Jessica Hamrick, Jason Grout, Sylvain Corlay, Paul Ivanov, Safia Abdalla, and Carol Willing. Jupyter Notebooks—a publishing format for reproducible computational workflows. page 4, 2016. doi: 10.3233/978-1-61499-649-1-87.
- Michel Launey. The features of omnipredicativity in Classical Nahuatl. *STUF - Language Typology and Universals*, 57(1), January 2004. ISSN 2196-7148, 1867-8319. doi: 10/ghxnzc.
- Eva van Lier. Lexical flexibility in Oceanic languages. *Linguistic Typology*, 20(2), January 2016. ISSN 1430-0532, 1613-415X. doi: 10/gdvjwt.

SPEAKERS IN THE CORPUS

sex	age	age_group	speaker
F	5	child	Anici
F	5	child	Selina
F	15	child	Kati
F	60	elder	Limeten
F	70	elder	Rosie
M	5	child	Enfant_2
M	10	child	Epel
M	10	child	John-Hari
M	30	adult	Aven
M	35	adult	Filip
M	35	adult	Kalbatik
M	40	adult	Bongman
M	50	adult	Dale
M	60	elder	Luan
M	60	elder	Sumbles
M	70	elder	Bogalgal
M	80	elder	Bongmeme
M	80	elder	Salben

... the relevance of the model:

T16.t2013.96 - 97

- Ara=tog-i van dry taid ili ara=kal: «Tck.»
- Ara=wuc moq-i, ara=lig plan avyn dry kal: «E

Aven then restarts with direct discours.