

De la collecte de données linguistiques à la création de ressources : coordonner les enjeux de la documentation d'une langue, le nisvai (Vanuatu)

Jocelyn Aznar



July 10, 2015

La problématique

Développer des ressources pour les différents acteurs intéressés ou participant à la documentation de la langue.

Les acteurs

- ▶ La communauté d'accueil du projet (et Le Centre Culturel du Vanuatu),
- ▶ La communauté des linguistes,
- ▶ Moi.

Mais il m'est impossible de développer des ressources «uniques» pour chacun de ces acteurs !

→ Ma solution a été d'adapter la forme de restitution suivant les acteurs.

Sommaire

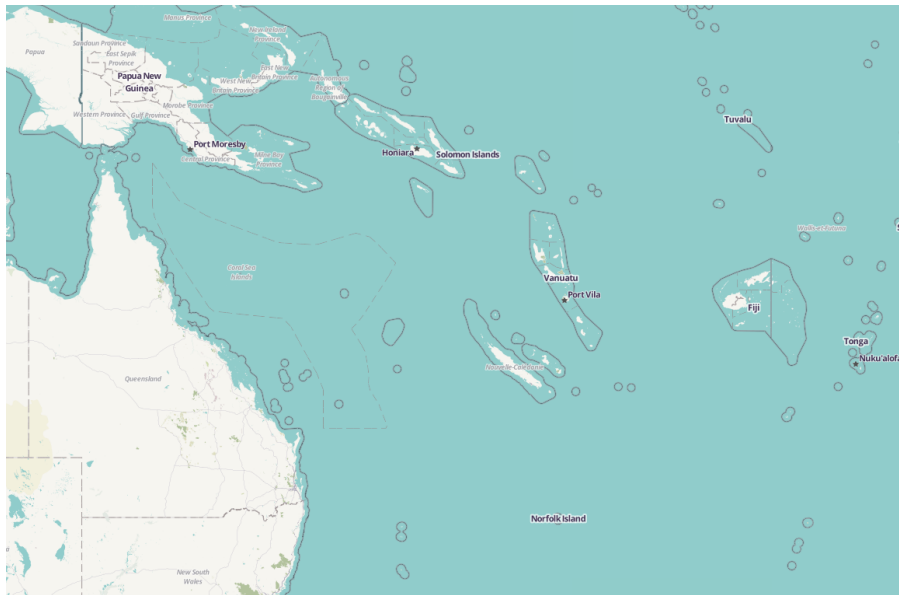
Introduction

La collecte des données

L'annotation du corpus

Une restitution fonctionnelle et différenciée

Vanuatu



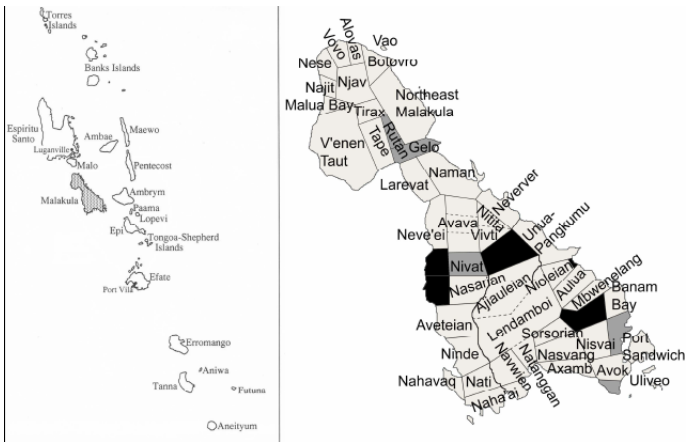


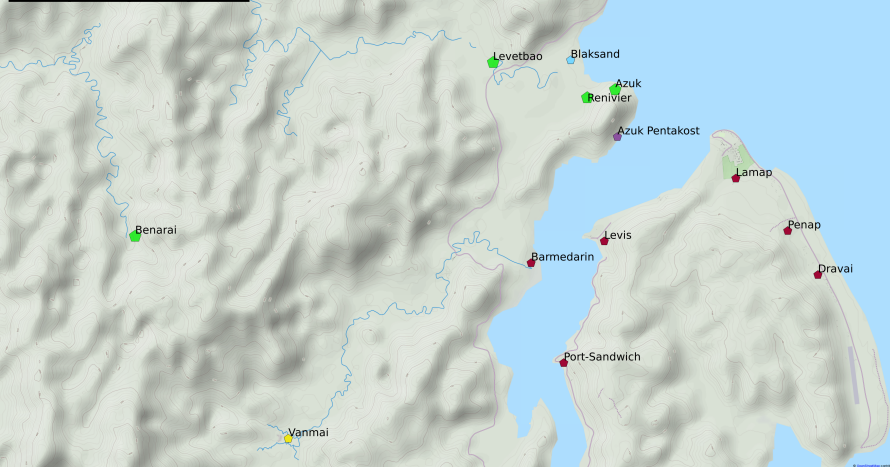
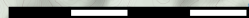
Figure: Les langues de Malekula, Source : Terry Crowley et John Lynch

Légende

Langues

- | | |
|---|---|
|  langue nisvaie |  langue de Port-Sandwich |
|  langue de Burbar |  langue d'Ambrym 1 |
|  langue de Pentekost |  langue d'Ambrym 2 |
| |  langue inconnue |

0 1 2 3 4 km



La collecte des données

Des ateliers pour une collecte des données ?

- ▶ Participation bénévole,
- ▶ Échec des ateliers,
- ▶ Enregistrements individuels (ou en «famille»),
- ▶ Conception locale du droit de transmission des récits,
- ▶ Corpus de textes organisés suivant des genres de discours émiqques.

Les données : des textes narratifs oraux

- ▶ Associable à des personnes,
- ▶ Marqué au début et à la fin,
- ▶ Possède un sens localement.

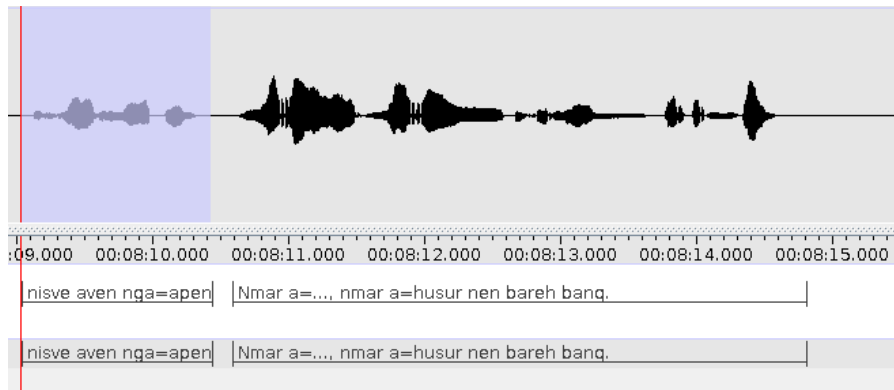
L'annotation du corpus

Quelques données sur le corpus de textes nisvaies

Des chiffres

- ▶ 14 textes annotés,
- ▶ 435 lexèmes différents,
- ▶ 1072 énoncés,
- ▶ durée d'environ 1h30,
- ▶ uniquement des histoires publiques,
- ▶ 10 orateurs différents.

La segmentation



Segmentation consiste à décomposer le flux continu de la parole en «énoncés» suivant les pauses des orateurs.

→ parcellisation des données tendance à perdre la cohérence et les marques de cohésion non-analytique.

No	Type1 : Transcription	Type2 : Reference	Type3 : Textualite	Type4 : Morphologie	Type5 : Traduction	Type6 : Note
130	Nga=kur-i nga=sah, nga=sah a-meh len bareh bang.	T11-2011.22	péripétie_3; LN_S	3SG=faire-INTR 3SG=grimper; 3SG=grimper TR.P-haut INST tronc banian.	Ce qui fait qu'il grimpe, Il grimpe en haut du banian.	
131	Nga=van [toux],	T11-2011.23	péripétie_4;	3SG=aller,	Il va.	
132	nisve aven nga=apen ?	T11-2011.24	péripétie_4; question rhétorique]	quoi REL.DEF 3SG=arriver ?	Qu'est-ce qui se passe ?	bislama:apen
133	Nmar a=..., nmar a=husur nen bareh bang.	T11-2011.25	péripétie_4;	TR.NM-serpent 3SG=suivre tronc banian.	Un serpent suit le tronc du banian.	
134	Nga=cumbrah, nisve a=kur-i ? Nga=ndras nga=cumbul.	T11-2011.26	péripétie_4; question rhétorique	3SG=se lever, quoi 3SG=faire-INTR ? 3SG=ne_pas_pouvoir 3SG=descendre.	Il se lève, qu'est-ce qui se passe ? Il ne peut pas descendre.	
135	Nga=cumbrah nga=kai.	T11-2011.27	péripétie_4;	3SG=se lever 3SG=chanter.	Il se lève puis chante.	
136	Nga=kai len nambé i-ang :	T11-2011.28	péripétie_4;	3SG=chante INST chant INTR-DEI.P	... cette chanson :	
137	Ofe ita je	T11-2011.29	chant;	MUS père MUS	Oh mon père,	
138	Kimaroh	T11-2011.30	chant;	IRR.1SG-etre,	si j'étais rester,	la langue est peut-être le nasvang. À noter que la répétition du chant dans la langue n'est pas exacte, si l'on prend comme référence Sumbles qui chante aussi la chanson, et qui connait cette langue.

Une restitution fonctionnelle et différenciée

Des scripts ad-hoc

- ▶ ElanToMysql.py → intégrer les données XML dans la base de données,
- ▶ gen_de_livret.py → transformer les annotations XML en fichier LaTeX,
- ▶ lexique nisvai → interface web au corpus.

cul la... navlun brav... nambrav aLivenbumbrao vi **bar** ili mecing bul a vi han. *tu fais que les poils du ventre de Livembumbrao vont frotter contre le tronc et ensuite tomberont par terre.* T12.2013.106

– **bare** tete

... talb len ni ara tams nen napenrevur ili len van **baren** a bubu pal. *qui ont porté le bois, ils tuent la vieille avec, sa tête est complètement fracassée.* T1.2013.139

... long swani len **baren**, nga hav a dri svoh cumbul. *... le cache dans sa tête, ensuite il saute.* T2.2013.45

– **bareh** tronc

Nga mestem, nen, ninqav sen nga roh len ameh len **bareh** banq. *Il manque, sa flèche reste en haut dans le tronc du banian.* T11.2011.19

Nmar a ... nmar a husur nen **bareh** nbanq. *Un serpent suit le tronc du banian.* T11.2011.25

Tamata acumbon ni nga husuh buluk van ili dri husuh nsal nen buluk nga van, bul nahemac bumi nahemac aven nga roh **bareh** bang ang. *Tamata seul suit la vache et suit le chemin que la vache a pris mais c'est le démon, le démon qui est sous le banian.* T1.2013.41

– **barih** tronc

Husuh hoi ili qam tvuk van dri lsi nalen nga tvuk **barih** nbanq, ale nga kai tasi : *Ils suivent la rivière, remontent en courant et voient les traces de pas qui montent vers le banian. L'enfant chante à nouveau :* T10.2011.31

nambur, *Il chasse le rat, le rat court et rentre dans un trou.* T13.2011.50

– **blah** atteindre

Bul nga sba **blah** nen naur bus... naur bus ili drembi. *Mais ils n'ont pas atteint le rivage encore.* T5.2013.45

– **blahe** atteindre

rohbai binbat vi **blahe**, bul nga vu nga sah bai. *il n'aurait rien, donc c'est bon qu'il grimpe.* T10.2011.16
Nga heli nga heli heli heli van dri **blahe** nga svoh, *Il creuse, il creuse et l'atteint, il (le rat) saute,* T13.2011.54

– **blav** long

Ngai meni nga sba **blav** ngai mutmut, nga cumbul. *Cela seulement, ce n'est pas long, elle est courte, elle se termine.* T3.2013.71

– **bo** odeur

a tongtong **bon**, *il sent les odeurs,* T12.2013.55

– **bong** jour

Ili ari ling **bong** aven. *Ensuite ils s'accordent sur un moment.* T1.2013.31

Bong ckai, *Un jour,* T13.2011.17

Nga kandr **bong** ckai, nanang, *Il y a un jour,* T8.2013.1

– **brah** avoir peur

Kik qanili aliroh ili a **brah** : «Be nga **brah**. *il parle comme cela et la femme a peur* : «N'aie pas peur. T4.2013.45

8 Ambat druan Livembrumbrao, Bongmeme Ngelu nga sihi Ambat et Livembrumbrao, par Bongmeme Ngelu

Nabong aven,¹ Ambat,² Welahlah,³ Sumbiehlu,⁴ Une fois,¹ Bat,² Welahlah,³ Sumbjehlu,⁴ Keskisvas,⁵ ara tuv roh van, ara kalik : «Dari, Keskisvas,⁵ Ils sont là et ils disent : «Allons-y, dara van van vani urer.»⁶ on va marcher jusque là bas.»⁶

Ara ... ara mur van, ara van van van van van acundr.⁷ Ara vanvan, aWelahlah a tait.⁸ «Ka ambat,⁹ Amdri ni mdra van, hana na rohroh urang ni.¹⁰ bena van.¹¹ Na tait monq.¹² Makamb, benqa rohroh, qa tong napen... nahevu ckai han nren ari nga bulah ili, nga han haing monqi.»¹³ Nga kal : «Awa, na roh suswah, be na roh lihai... percang len sal be na roh suswah.»¹⁴

Nga rohroh van ili ngai va ... va roh suswah.¹⁵ Be nahemac ili qa nambwol... Livembumbrao tongtong nambwon nren a qan naur lipah dri tong nambwon nambwas. ¹⁶ Tongtong bwon va van va van bulrohni.¹⁷ Van bul : «Koh¹⁸ A mhambong va ... van qa minmin va roh urha?¹⁹ Ko, na... mandra mai, mandra mai vi ureh, be na vanvan na tong na samb.²⁰ Han bul rohroh urang Ambat ara van. Oh qa min bai, dara van viim.²¹ ²² Dara van viim bai, dara va kan bai.²³ dara van Ambat ara mai ili, mramcu lulu.»²⁴ Nga kal : «Awa, hana na ngot. Na rohroh urang ni. Nga kal va puc ahnaw. Amel puc haing, dara van viim drongdong kan dris hani.»²⁵ AWelahlah ngur ngot, nga paci nga sungi len nandr.²⁶ Talbi.²⁷

Ils vont, ils vont très loin.⁷ Alors qu'ils vont, Welahlah est fatigué.⁸ «Eh Bat,⁹ Vous autres vous y aller, moi je reste ici.¹⁰ Je ne vais pas.¹¹ Je suis déjà fatigué.¹² Si tu restes, fais attention à une vieille femme qui mange les personnes, si elle te trouve, elle te mange.»¹³ Il dit : «Non, je reste caché, je ne serai pas dans la for... près du chemin, je serai caché.»¹⁴

Il reste un peu et va se cacher.¹⁵ Mais le démon, Livembumbrao, sent l'odeur des hommes comme les chiens sentent l'odeur des cochons.¹⁶ Il sent, il sent et il les trouve.¹⁷ Et il dit : Oh,¹⁸ Ah, mon grand-père, qu'est-ce qui t'amène ici ?¹⁹ Ah, nous venions, nous venions là-bas, mais je marchais puis je me suis senti mal.²⁰ Je suis alors resté ici, et Ambat et les autres sont partis. Oh approche, allons à la maison.²¹ ²² On va à la maison et on mange.²³ On y va, quand Ambat et les autres viendront, vous rentrerez chez vous. »²⁴ Il dit : «Non, je ne veux pas, je reste ici. Il dit : tu me mens. Mais non je ne te mens pas, on va à la maison chercher à manger et manger.»²⁵ Welahlah ne veut toujours pas, il le porte et le met dans son sac.²⁶ Le met sur son épaule.²⁷

Une ressource en ligne

Corpus de textes narratifs nisvais

Recherche dans les textes :

Vous devez choisir un résultat apparaissant dans la liste proposée par l'interface.

nahemac démon

bulbi mra=tvuk van vi bareh nbaq i-sao seh nahemac aven nga=roh-roh bareh

? 2PL=monter aller DIR tronc ARBRE ?-DEI.E COO.CSQ démon DEF 3SG=etre-RED tronc

mais vous n'allez pas là-haut au banyan car un démon se trouve là-bas.»

Ale Tamata cumbo-n ni nga=roc buluk aven van. Bul buluk nen nahe

COO.TMP PRENOM seul-3SG INT 3SG=chasser vache COO.DEF aller COO.OP vache DEI.?? dém

Puis Tamata, seul, chasse la vache qui est partie mais cette vache, c'est un démon.

Conclusion

- ▶ Processus de documentation progressif d'une langue orale en cours de description,
- ▶ Le texte comme unité d'observation,
- ▶ Une source unique → des ressources différenciées,
- ▶ Une limite majeure : si la forme est adaptée, le contenu ne l'est pas.

Merci pour votre attention.

- ▶ courriel
`contact@jocelynaznar.eu`
- ▶ Les ressources papiers
`http://37.187.224.23/trela2015/`
- ▶ L'interface Web
`37.187.224.23/lexique_nisvai/form_lexique.html`