

Sparse Transcription Rethinking Oral Language Processing

Steven Bird, Charles Darwin University steven.bird@cdu.edu.au (paper for this talk available on request)



Why we transcribe



It is easy to capture a large amount of audio

We want to get text into the linguistics or NLP pipelines

But: transcription bottleneck

The tools



) 🔛 🗙 🗠 🔉 😓 🖕 🦛 🛶 🐗 🥘 🗞 📴 🚰 🎂 🗐 🗄 English & Texts Text near T Title Title ordanc Show All List C John 2:17-22 -Address, section A version 2.way Info Baseline Gloss Analyze Tagging Print View Text Chart Edt W John 2:12-17 1 Word John 2:1-12 Guaj xton Morphemes John 1:43-51 guaj xton Lex. Entries John 1:35-42 Lex Gloss guaj³² xton³² John 1:19-34 okay · · · John 1:1-18 Free Eng Okay John (Title) Spn El zorro y las tunas Lit El sol y la luna, version 1 rihaan cuan <mark>ne</mark> ranga' 2 Word Soj me ranga' Address, section A version ne Morphemes SOI me ranga' rihaan 888 ne ranga' ne 2.wav *** Lex. Entries zoj³ me³ 3 ranga^{'3} riaan³²1 ni³+dial. var. ranga'³1 ni³+dial, var. Acts 9:35-43 *** Lex. Gloss you (pl) be buenos dias to negative buenos dias negative Acts 9:31-35 Acts 9:26-31 Free Eng You who are in the morning Acts 9:22-25 Spr Lit Acts 9:19-22 Acts 9:1-19 3 Word Soj me auun xta' guii rihaan cuan ne guun xta' guiii xico Acts 8:25-40 *** guii rihaan *** ne g- uun *** Morphemes SOj me g- uun me³ 3 c-1 uun³ ra⁴ *** gwii³ riaan³²1 *** ni³+dial.var. c-1 uun³ ra⁴ ----Acts 8:1-25 Lex. Entries zoj3 Lex. Gloss you (pl) *** *** *** *** Acts 7:53-60 be com want day to negative com want Acts 7:2-53 ne Acts 7:1 ne sts Acts 6:7-15 ni³+dial. var. negative - Acts 6-1-7

Elan: Glossing with no lexicon!

FLEx: Transcription with no speech

Edit View Data Insert Format Tools Parser Window Help

Computational approaches NO TWO THE SAME

Citation Hanke and Bird (2013) Sloetjes, Stehouwer, and Drude (2013)Anastasopoulos and Chiang 00 (2017)Rogers (2010) Do et al. (2016) Chung et al. (2019) 0 Duong et al. (2016) Adams (2017) Besacier, Zhou, and Gao (2006) Adams et al. (2016b) Foley et al. (2018) Anastasopoulos and Chiang (2018)Stahlberg et al. (2015) Zanon Boito et al. (2017) Godard et al. (2016) Anastasopoulos et al. (2017) Godard et al. (2018b) Duong et al. (2016) Besacier, Zhou, and Gao (2006) Anastasopoulos, Chiang, and Duong (2016) Adams et al. (2015) Stahlberg et al. (2016) Keyword spotting Objective



What is transcription?



Kunwinjku: polysynthetic language, 2000 speakers

kabirrimalaydjarrkkunjdjangkan

birri- malay- djarrk- kunj- djangka- n we.incl morning together kangaroo hunt non.past

let's all hunt kangaroo together in the morning

Source

- 1. Record
- 2. Respeak = oral transcription (now)
- 3. Transcribe (sometime in the future)

BALANG NABIRDBIRD GUNBALANYA OCTOBER 2018

Oral Transcription (interactive)

ku-bolk-nahnah-ni

birri-bolk-nahnah-ni

birri-bolk-nahna-ni they-country-look.after-imperf

Do enough so I can transcribe later

BANGARDI MARALNGURRA GUNBALANYA OCTOBER 2019 birri-h-ni birri-bolk-nah-na-ni ku-mekke kun-red they-BE they-look-after that country

What ngandibektan Ngay Balang Nabirdbird Wanjh ngayaw mulewan bin nahne bin kahbindi Kubodnie ngarduk nahne ngurrihnan Ublobbo kure nahkohbang ngardunk konkumo marney dydneg Buladjang. Nahne walen. Koroko birrihni mawalsmawah ngabenlyine bioriyatiyatiwurd dja birriwernwarse dja mak yæyow mawahmawah ngabenyine Yoh ngalolaard n ugadoerre kurnelde birrihdi nay ung ki navahnavah nak navah kabiriyung ki bolk navahnavah konoko binihni birrinahnani

OCT 3 2019 Jorenials

Mah ngandibektan Ngay Balang Nabirdbird Wanjh ngayawmulewan bin nahne bin, kahbindi "Kubodwe ngarakuk nahne ngurihnan dolabba kwe nahkahbang ngarakuk konkumo morneydydnu Buladjang. Nahne walem. Konoka birrihni Mawal nawah. ngabengi me bioriyadyafarward dja birri wennare dja nak yasan nawahmarah ngakengi me Yah ngabara ngabengi kunaka birrihdi nagungki nawah nawah nak namah nakiriyu ng ki nawah nawah konoko birrihni birrinewani

Kore Konred Kore Buladjang kaldi wanji ngurina ngurina kali Nahne kali kunekte kore birihuram naney nangy nguni kunekte kore nabang kalyo nangy nguni guni nak kunet birdidahneng unang nguni bindida kore nabang kalyo naware nametike yarika kawarim nang i naware nametike yarika kawarim nang yalyawaring yunan kakumane mungun ungu

navare namen nave kokulja birilje yahyawunini, yinan, kokumaa mungui mungui Wanja birihuan minekke biriljalet dawu birdinanebebmeng navekasiare Bulajang Wanja biri karungi yinan cayine maga bidbuni biri ilatemi kumande kundulk Kalawan yinan kayine 3 navu koho ko bagi birikang birikawagi kurabaaka, mankindjek, ngarbek biriyamani

Kundder Kaniwakawakaran ngawakaran nane mayh birnigawan kadoroki hanjih birniwan yimen birni bolkwarch Warchong Birnikanag nandke mayh birnikolkbengbom baleh wani birnikanag nandke mayh birnikolkbengbom Kalibindi kuladik kubodane ngancuk wanih kumekke yimen bolkrokang manjih bolkbakabakmeng bakrokang wanjih bolk bakabakmeng kuji b birniyimeni boliri jubaneni bakyik birniyimerneni ku kuta

Deluk keh bining kahadikahang balkbakabakaneng Kun viander lakabakaneng viandi belediplani viano dinasak balkakanukan balah bining bini ilakunging barah

Bernh Nawu bini biniyibmeng ^{kare}kanjadji Kab bin kabbindi kuni kare binire mandjiandjin bakenh

Man labball kadi kunekke yinan kayi an watefall manu kunekke solk bahadearikmen

Wanijh birrigineginerring wanijh nemekke kunih kuni Man benglanikbon 8 rowk kunwardhe bengharibon birniginerran birriginegine birrigineunguh

Mah ngadibekkan ngaye balang nabirdbird wanjh ngayawmulewan nahne bim kahbimdi nawu kubodme ngardduk nahne ngurrihnan dolobbo kure nahkohbanj ngardduk kornkumo nganmarneyolyolmeng Buladjang nahne walem dja korroko birrihni mawahmawah ngabenyime birriyahyahwurd dja birriwernwarre dja mak yayaw mawahmawah ngabenyime yoh ngabbard kornkumo ngadberre kumekke birrihdi nayungki nawu mawahmawah mak nawu mawah kabirriyungki ?? mawahmawah korroko birrihni birribolknahnani kore kunred kore Buladjang kahdi wanjh ngurrina kahbimdi nahne ngurrina kahbimdi kumekke kore birrihwam nahyungki birrire mayh birriyawani kore bindidahmeng bindimarneyimeng mak yuwn mak



wanjh ngayawmulewan nahne bim kahbimdi nawu kubodme ngardduk nahne ngurrihnan dolobbo kure nahkohbanj ngardduk kornkumo nganmarneyolyolmeng Buladjang nahne walem dja korroko birrihni mawahmawah ngabenyime birriyahyahwurd dja birriwernwarre dja mak yayaw mawahmawah ngabenyime yoh ngabbard kornkumo ngadberre kumekke birrihdi nayungki nawu mawahmawah mak nawu mawah kabirriyungki ?? mawahmawah korroko birrihni birribolknahnani kore kunred kore Buladjang kahdi wanjh ngurrina kahbimdi nahne ngurrina kahbimdi kumekke kore birrihwam nahyungki birrire mayh birriyawani kore bindidahmeng bindimarnevimeng mak yuwn mak ngurrire kumekke bukka kore nabang kahyo nawarre namekke yarrkka kawurrhme wanjh yahuawurrinj nawu yiman nawu birridjale munguhmungu wanjh birrihwam manekke birridjalam nawu bindimarnebebmeng namekke nawarre Buladjang wanjh birrikarungiviman kavime mavh bidbuni birridelmi kunwardde kundulk kalawan viman kavime dia mak nawu kohbohkohbani yiman birrikang birrikarungi karlabarda, mankindjek, ngarrbek birriyawani kunukka yarrkka karriwakawakwan nawu mayh birriyawani kadberre ki wanjh birriwam viman birribolkwarrehwarrewong birriwarrewarrewong irribolkbengbom

Source

mah~ok

dolobbo~bark

birri~*they*

nahne~this

dja~and

mawah~ancestor

BALANG NABIRDBIRD GUNBALANYA OCTOBER 2018

Interactive Transcription

Play a recording from the computer

Pause

Speak what we are hearing

KAMARRANG GUYMALA KABULWARNAMYO OCTOBER 2018



What's going on?

- We recognise repeated forms, in the midst of unrecognised material
- There is always unrecognised material
- We can only skip it (wastes time to try to transcribe it)
 - noise in the signal (ambient, human)
 - disfluency, speech impediments
 - unknown vocabulary (incl loanwords)
- More examples of why we want to transcribe words, not phones...

Transcribing words

- 1. kayadirri ~ ka-birri-yaw-dirri
- 2. berre ~ bedberre
- 3. mahne ngalengman ngan-bedde
- 4. ka-bourk-mang ~ ka-borurrk-mang ~ ka-bo-durrk-mang
 - 5. kadiakodjuke ~ konhda ka-bandi-yaw-kodj-djuhke



English example: d'ya d'ya see? DO WE WRITE WORDS OR PHONES?

- 1. dog jog si enables further analysis of the text
- 2. czaczasi enables accurate phone recognition

`field linguists [should modify] their [transcription] practice so as to assist the task of machine learning' – Seifart et al 2018

`linguists should aim for exhaustive transcriptions that are faithful to the audio ... mismatches result in high error rates down the line' – Michaud et al 2018



We hear a form repeatedly, and add it to our list, with a canonical representation Many of these speech tokens will be significantly reduced Transcription = pairing a locus of speech with an entry in a proto-lexicon

This is not new

"I listened to stories and wrote down words... my glossary is really growing"

– Franz Boas (quoted in Sanjek 1990)



FRANZ BOAS 1858-1942

"No matter how careful I think I am being with my transcriptions, from the very first text to the very last, for every language that I have ever studied in the field, I have had to <u>re-transcribe my earliest texts</u> in the light of new analyses that have come to light by the time I got to my later texts... You can probably expect to be transcribing and re-transcribing your texts until you get to the final stages of your linguistic analysis and write-up." (Crowley 2007)

"... a transcription, whatever the type, is always the result of an analysis or classification of speech material. <u>Far from being the reality itself, transcription is an abstraction from it</u>. In practice this point is often overlooked, with the result that transcriptions are taken to be the actual phonetic 'data'." (Cucchiarini 1993)

Those who deal with the spoken word... seem to regard

phonography as little more than a device for moving the scene of

alphabetic notation from the field interview to the solitude of an

office... The real analysis begins only after a document of altogether

pre-phonographic characteristics has been produced... The

alphabet continues to be seen as an utterly neutral, passive, and contentless vehicle – Tedlock 1983

In spite of this...

Documentary workflow: transcriptions = data

The importance of the transcript resides in the fact that for most analytical procedures it is the transcript and not the original recording which serves as the basis for further analyses – Himmelmann 2006

For the scientific documentation of a language it would suffice to render all recordings utterance by utterance in a phonetic transcription with a translation – Mosel 2006

Transcription? phonetic vs IGT

alwanjukgorogoalgo?banbaribajimiwoubaleianjaibabuni

al-waŋɟuk gɔdɔgɔ al-gɔ?baŋ ba-di ba-jim-i. woh, ba-ɹe-i man-ɟai, ba-bu-ni II-emu before II-old.person 3P-beP 3P-say-PI yes 3P-go-PI III-cane.grass 3/3P-hit-PI

Workflow: phonetic transcription \rightarrow glossed text

NLP pipeline: transcriptions = data



- 1. automatic phone transcription
- 2. automatic word segmentation
- 3. the rest of the pipeline...

Word segmentation? BAKED INTO THE DATA!

tempm ~ ten pm'ten pin' (homorganic nasal assimilation) $hæd\underline{s}i \sim hæd ji$ 'had your' (palatalisation)tentsents ~ ten sents'ten cents' (coarticulation of nasal and fricative)lbrænd ~ lb ænd'law and (order)' (intrusive 'r')

Who can do this? ANSWER: ONLY (COMPUTATIONAL) LINGUISTS

phonetic transcription

IGT

automatic phone transcription

automatic word segmentation

TITLE) THE BLIND WEMAN AND HER SEN" 31/05/2. Velaliki Vena Koa ei gipala O Alo gozoba vena makokisa gipala isa minasing Long timengo woman one son both Stayed. lived. © Alo gozobą vena makołaśą gipalą isą minąsiny Long timego woman one von both Stayed lived. e Menipo zoliha venala zegipa getamiasko hili**ti**. Father nityet wife baby bern dieg lived. 3 Zegipa getoake gizipa otoko itina. baby born looked after grow up. Denipo zoliha venala zegipa getamiarsko hilita. Father notget wife baby born dieg e Mota litasko napadoake iza nama peletikang 6 Veny quickly grews pig birds killed une S Belahina gipopa otoko vina. Hismother lookadafar went Vera manikeko umuhilina tana. vife identify liked got beautiful married 3 Zegipa getoake gizopa otoko itina. baby born looked after grew up. Aununar aits giziki ikasina house separate build slight Nuwaka nuwaka nosaneta titiki vika asina. ajkinom ajkinom food browht war usa to Dery quickly grews pig birds Killed me lexical Blzelahing gjjopa otoko vina His mother looked after went identifier





no segmentation we can always discover elided material between words

Collaborative workflow PhD project of Éric Le Ferrand



- 1. for each phrase:
- 2. verify forms automatically recognised in previous iteration
- 3. tag with new lexical items that we can confidently identify
 - a. speak a form that was not identified
 - b. automatically locate it in the phrase
 - c. elicit translation, add to lexicon
- 4. retrain word models

Scaling up









1)#1+1)===	
\$14192-1	
110-Hz- 41081-11-111-11	
- 1100 D - 110 Ho - 1	
Helsifeter form these	
He	
eren der Herende	
to for 1100 to 100 fto	
a 110 fla \$2001- 11- 111	

TRANSI ATION



(c) Tapered Corpus: an open-ended speech collection, some translations, fewer transcriptions, cf Fig. 2

(a) Low Resource Scenario, with enough data to train a basic speech recogniser

(b) Zero Resource Scenario: unsupervised word spotting by co-indexing and labelling recurring forms

From protolexicon to lexicon

New PhD project!

Protolexicon contains repeated forms. We need to analyse it, splitting and merging...

Only then will our transcriptions be a `nice polite sequence of morphemes' (Bender 2019)

Construct the actual lexicon

Evaluation: "Lexeme error rate"

Dismantling the transcription bottleneck

If transcriptions = data

- 1. want a surrogate for the signal: phonetic transcription
- 2. must transcribe everything
- 3. must transcribe as the first step

If transcriptions = observation

- 1. want any useful annotation of the signal: includes "lexical" transcription
- 2. we transcribe what we can observe (always provisional, always more audio)
- 3. we can translate first

This is a return to orthodoxy: The Tapered Corpus



- (1) **Core Corpus**: a central body of data, the basis of a complete linguistic analysis;
- (2) **Indexed Corpus**: equipped with a complete lexicon, an indexed list of morphemes with glosses and morphological classifications;
- (3) **Transcribed Corpus**: transcriptions prepared as soon as possible after the recording to reduce the frustrations of cold notes;
- (4) **Translated Corpus**: not transcribed but translated into some familiar language, with indications of the social contexts;
- (5) Raw Corpus: unprocessed recordings.
- The quantities of data at each level follow a power law, based on the amount of curation they require (after Twaddell 1954, Samarin 1967)
- Translation precedes transcription (capturing meaning is more urgent than re-representing our data)

Three approaches to design

How does an outsider encourage people to keep their language strong?

- 1. capture (is it effective?)
- 2. address a cause of language shift: low prestige, relevance
- 3. address another cause: functionality

LanguageParty.org



Notre Language Party!

UNIVERSITÉ PERPIGNAN VIA DOMITIA Berbère, Mongol, Kanak, Arménien, ... Contes dans les langues d'origine avec traduction mardi 26 nov à 18h30 à la bibliothèque

Oral Language Learning APPROPRIATING SIMPLE APPS



(a) Learning terms of address by capturing selfies and recording a short bilingual dialogue

÷	Files are stored in: /storage/emulated/0/Kunw	okCl
	1 smoke in my eyes	00:12
	September 2, 2018 at 4:56 PM	375 kB wav
	1 take a little bit	00:09
	September 2, 2018 at 5:55 PM	782 kB wav
	1 They told me theyre	00:15
	coming today September 2, 2018 at 4:56 PM	472 kB wav
	1 Was coming to see you	00:09
	September 2, 2018 at 4:55 PM	290 kB wav
	1 white skin black skin	00:06
	September 2, 2018 at 4:56 PM	207 kB wav
	2 cheeky yam	00:19
	September 2, 2018 at 4:55 PM	615 kB wav
	2 cold and warm water	00:06
	September 2, 2018 at 4:55 PM	205 kB wav

(b) Learning key phrases by capturing segments where a term is spoken and translated



Back Forward

(c) Learning linked to places, where multiple people describe the same set of places

Conclusion 1: The Status Quo

In NLP, we extend our text-based methods to handle speech input by adding a speech-to-text component

Therefore, for unwritten languages: convert speech data to text data and continue as before

The existence of automatic speech recognition provides proof of concept.

But, these are not "unwritten languages", but ...

Conclusion 2: Problems with status quo

- over-values transcriptions, treating them as data when they are nothing more than contingent observations (`*Premature systematisation keeps nature's surprises hidden'* – Lenat and Feigenbaum 1987)
- 2. **under-values words**, treating them as the byproduct of boundary detection when they are meaningful units which often overlap in the speech stream
- 3. **trivialises the role of linguists** to phone transcription when they can conduct complex workflows involving iteration and interaction
- 4. excludes the speech community, the workforce, and the main beneficiary (NB for these people, FAIR != fair)

Conclusion 3: Transcription as Observation

transcription = data: transcribe phones / fully / first

transcription = observation:

- map locations in the speech stream to an inventory of meaningful units
- generalises over dense and sparse transcription
- transcription = discover repeated meaningful units (regardless of whether they are canonical)

Conclusion 4: Have we addressed the bottleneck?

Don't waste scarce resources on unimportant tasks!

- phone transcription is extremely time consuming for spontaneous speech in the presence of coarticulation and disfluency
- much of the phonetic detail is not necessary as long as we are identifying words

Instead, allocate resources to the central acts of transcription:

- identify meaningful units without deciding on their formal status (as morphemes, words, or multiword expressions)
- identify meaningful units without baking-in boundaries (then train word-spotters)
- allocate effort to the units of interest (ie words with meanings), to improve our ability to identify oral texts for closer attention (including dense transcription)

Conclusion 5: New promises of scalability

- 1. it is open to participation by local speakers, given their superior ability to identify meaningful units in the speech stream, even in the presence of noise
- 2. it faces the Zipfian distribution of words; word-spotting enables us to allocate effort according to decreasing frequency, and to the topics / texts of interest
- it treats each additional resource as auxiliary information, i.e. further supervision to help annotate the signal; this necessary flexibility in the face of diverse language situations, different constellations of data and skills