

# Digital Databases for Linguistics

## CSUN Open Access: A Day of Data, 10.25.16

October 21, 2016

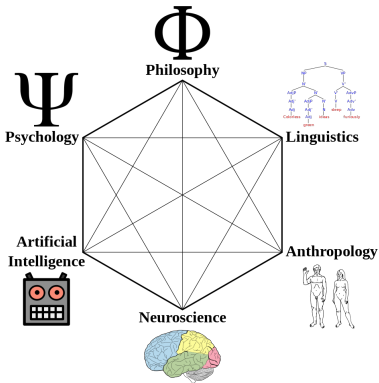
# Outline

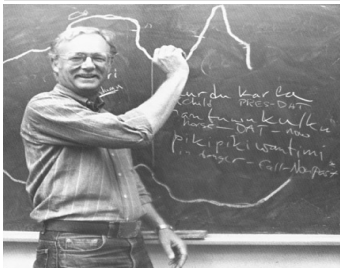
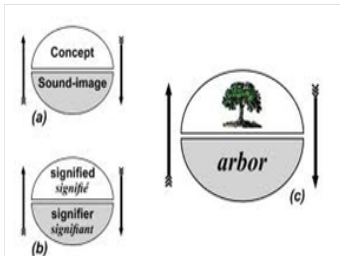
- 1 Roots of Formal Linguistics
- 2 Project: Valency Change in Hawaiian
- 3 Teaching

# Outline

- 1 Roots of Formal Linguistics
- 2 Project: Valency Change in Hawaiian
- 3 Teaching

# Linguistics and Cognitive Science





## Antecedents



$$e(p) = e_v(p)$$

$$e(\perp) = 0$$

$$e(\top) = 1$$

$$e(A \otimes B) = e(A) * e(B)$$

$$e(A \rightarrow B) = e(A) \Rightarrow e(B)$$

$$e(A \wedge B) = e(A) \wedge e(B)$$

$$e(A \vee B) = e(A) \vee e(B)$$

$$e(A \leftrightarrow B) = e(A) \Leftrightarrow e(B)$$

$$e(\neg A) = e(A) \Rightarrow 0$$

# Outline

- ① Roots of Formal Linguistics
- ② Project: Valency Change in Hawaiian
- ③ Teaching

# Valency

- (1) a. bath - bathe  
b. tooth - teethe

- (2) a. haole  
*white person*  
b. ho'ohaole  
*to act like a white  
person*

- (3) a. koa'e  
*tropicbird*  
b. ha'akoa'e  
*to act like the  
tropicbird*



## More examples

(4) sit - set, lie - lay

- (5) a. 'āpuka  
to cheat  
b. ha'āpuka  
to cause to cheat

- (6) a. 'ai  
to eat  
b. hō'ai  
to feed



## Puzzle

- (7) a. nui  
*large*
- b. ha'anui  
*to brag, exaggerate*
- c. ho'oha'anui  
*to cause to brag*

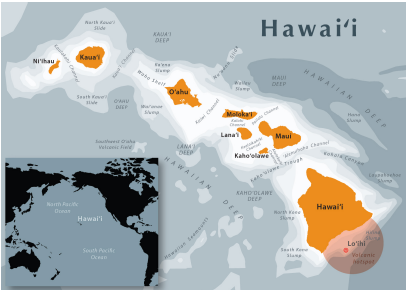
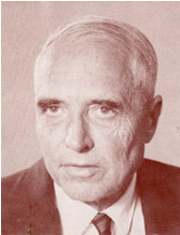
- (8) a. ko'o  
*brace, prop, pole*
- b. ho'oko'o  
*to prop with a pole*
- c. ha'ako'o  
*(same as ho'oko'o)*

# Prior Analysis

Roots of  
Formal  
Linguistics

Project:  
Valency  
Change in  
Hawaiian

Teaching



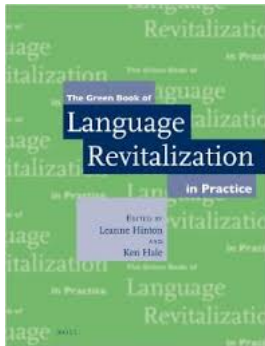
- 836 entries headed by *ho'o* & *ha'a*
- Coding for morphology, phonological alternation, interaction with reduplication, meaning, etc.

	A	B	C	D	E	F	G	H	I
1	ID	base	basepos	basedef	base notes	reduplication meaning	derived form	derdef	notes
49	M048	maa.ino.ino	redup of nvs		redup. of maaino 'cruelty'		ho'o maa.ino.ino	caus/sim; to torture; atrocity	
50	M049	maa.'ino.'ino	vi	to deface, spoil, ruin; defamed, defaced			ho'o maa.'ino.'ino	to defame, slander; deface	
51	M050	maa.lo	vs	furrowed, grooved, out in ridges; very thin, wasted			ho'o maa.lo	to furrow, cut in ridges, cause thinness.	
52	M051	maa.'yo	vt	to cut raggedly and unevenly, as cloth or hair; dented; chipped, as crockery			ho'o maa.'yo	caus/sim	
53	M052	maa.'lu.'lu	vi	at a distance, out of sight.			ho'o maa.'lu.'lu	to keep at a distance, avoid	
54	M053	maka	n	eye, eye of a needle, face; presence, sight; lens of a camera			Ho'okeesamaka	to be partial, show favoritism	note that the base in the h
55	M054	maka	n	Point, bud, protuberance; center of a flower; nipple, test; sharp edge or blade of an instrument; point of a fishhook;			ho'o maka	to begin, initiate, to appear, of a child's first tooth; to put forth buds;	
56	M055	maaka	nvt	mark, marker, blaze, target; to mark			ho'okilikii maaka	target shooting	note that the base in the h
57	M056	maka.'ee	vt	to look at with disfavor; to look askance.			ho'o maka.'ee	caus/sim	
58	M057	maka.'eo	vs	angry-appearing			ho'o maka.'eo	to look at with anger, to avoid looking at because of anger; not to	
59	M058	maa.kaha.kaha	nvi	clearing, as rain			ho'o maa.kaha.kaha	to show signs of clearing	
60	M059	maka.hehi	nvt	admiration, desire for; attractive; to admire; be entangled			ho'o maka.hehi	caus/sim	
61	M060	maka.hinu	n	bright face, cheerful look			ho'o maka.hinu	caus/sim; to conceal annoyance or anger by pretense of cheerfulness	
62	M061	maka.hou	n	beginning, new start			ho'o maka.hou	to begin again	
63	M062	maa.ka'i	nvt	police man, guard; to police, inspect, spy			ho'o maa.ka'i	to act as a policeman; to appoint or invest as a policeman	
64	M063	maa.ka'i.ka'i	nvi	to visit, the sights; to stroll, make a tour; spectator			ho'o maa.ka'i.ka'i	to take others on a visit; to show the sights; to escort	
65	M064	maka.kii	vt	to look at with hatred			ho'o maka.kii	to look at with hatred; to plan revenge or evil	
66	M065	maka.kiu	nvs	spy, detective, spying eye, watchful, vigilant, spying; to spy			ho'o maka.kiu	to spy, watch, reconnoiter	
67	M066	maaka	vc	to loosen, unbind, untie, liberate; to remain at a distance; to front			ho'o maaka	caus/sim	

- Qualitative Analysis, Descriptive & Inferential Statistics



## Open Data



# Outline

- ① Roots of Formal Linguistics
- ② Project: Valency Change in Hawaiian
- ③ Teaching

# Computation Linguistics



LING 455

Computational Linguistics

Day, time – time, Building tba



Instructor:

email:

Office:

Office hours:

tel:

**Course Description:** This course is an intensive seminar on computational data processing in linguistics. While computers have no minds of their own, researchers are increasingly able to harness computational methods to understand natural language, which is often 'messy' and unpredictable. The focus of this course is the creation of working computer code that can help us answer questions about language by accessing resources from basic corpora (e.g. structured narratives like fables or news stories) to social media. No specialized knowledge of computer science is necessary. The skills learned in this class will be useful for undertaking research in linguistics and the social sciences, as well as for performing data analytics on text corpora. Available for graduate credit.

## Course Objectives:

At the completion of this course, students will be able to:

1. Write simple but effective programs in Python
2. Analyze language data (e.g. text corpora) using Python
3. Understand how data structures and algorithms are used in NLP
4. Develop and answer language-related questions on the basis of data available on the internet

## Required Texts:

Bird, S., Klein, E., & Loper E. (2009). *Natural Language Processing with Python*. Sebastopol, CA: O'Reilly Media. ISBN-13: 978-0-596-51649-9

Available for free from the author, at [http://www.nltk.org/book\\_led/](http://www.nltk.org/book_led/). This can also be purchased as an ebook from O'Reilly (recommended, if purchasing) or in paper or ebook form from all the normal places.

Russell, M. (2014). *Mining the Social Web* (2nd ed.). Sebastopol, CA: O'Reilly Media.

ISBN-13: 978-1-449-36761-9 Note that we will use Ch.1 only

Chapter 1 is available for free from the publisher, at <http://bit.ly/135dHfs>. The entire book (which we will

# Goals

- Big (biggish?) data
- Digitized - easy to access
- Open
- Thank you!