

## 接頭辞の処理は困難か？ 類型論的少数派からの考察

浅尾仁彦

(ニューヨーク州立大学バッファロー校)

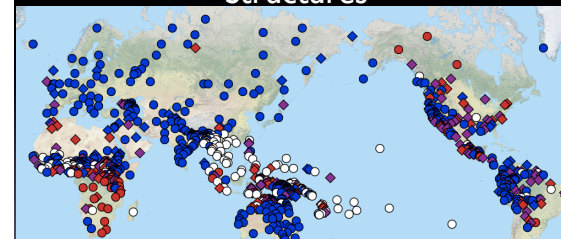
asaokitan@gmail.com

KLS39 (大阪大学) 2014.6.14

## 接辞の左右非対称性

- 典型的に接頭辞よりも接尾辞が多い  
(Sapir 1921, Greenberg 1963, Hawkins & Gilligan 1988, Bybee et al. 1990, Dryer 2013)

## World Atlas of Language Structures



赤: 接頭辞型 (152) 青: 接尾辞型 (529) (Dryer 2013)

## なぜ接頭辞は少ないのか？ 言語処理からの説明

- Greenberg 1957
- Cutler et al. 1985, Hawkins and Cutler 1988
- Hall 1988, 1992
- Himmelmann 2012
- Asao 2013

## Asao 2013

- 短い要素は曖昧性が高く、聞いたその場で特定するのが難しい
- [paip] という音を聞いた場合、pipe という語が用いられた可能性が非常に高いが、[s] の音だけで複数接辞 -s と決定することはできず、文脈の情報が必要

## Asao 2013

- p a<sup>↑</sup> p - s
- この時点で pipe という語が使われたことに確信を持つことができる
- s - p a<sup>↑</sup> p
- この時点で s という接辞が使われたとは確定できない

## “類型論的少数派”問題

- 「接頭辞の処理が困難」なのだとしたら、類型論的に少ない接頭辞タイプの言語が問題なく存在していることをどう説明するのか
- 類型論的傾向の「心理的説明」に常につきまとう問題

## “類型論的少数派”問題

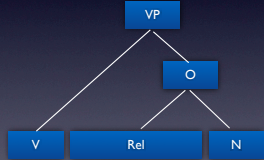
- Harris (2008) 「接中語 (endoclititic) のようなものが稀なのは、処理や獲得が困難だからではなく、単にそれを生じさせるような歴史的偶然が起こりにくいだけ」

## 本発表の仮説

- 類型論的傾向に対する「心理的」説明は有効である

## 関係節語順の例

- VO 言語で、RelN（関係節前置）語順は非常にまれ
- 中央埋め込みを生じるから (Kuno 1974; Hawkins 2002)



## 関係節語順の例

- 中国語は VO & RelN
- 中国語は処理が困難なのか？
- 中国語には主題化構文があり、これが中央埋め込みの困難を回避する目的で使われている可能性 (Matthews & Yeung 2001, Ching 2008)

## 関係節語順の例

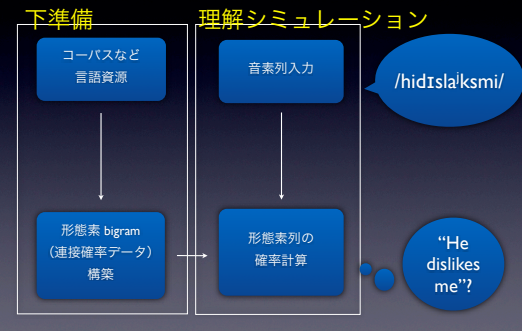
- 主題化による中央埋め込みの回避
  - Lei<sup>5</sup> zeon<sup>2</sup>bei<sup>6</sup> hou<sup>2</sup> [ting<sup>1</sup>jat<sup>6</sup> gong<sup>2</sup> go<sup>2</sup> di<sup>1</sup> je<sup>5</sup>] mei<sup>6</sup> aa<sup>3</sup>?
  - [Ting<sup>1</sup>jat<sup>6</sup> gong<sup>2</sup> go<sup>2</sup> di<sup>1</sup> je<sup>5</sup>] lei<sup>5</sup> zeon<sup>2</sup>bei<sup>6</sup> hou<sup>2</sup> \_\_ mei<sup>6</sup> aa<sup>3</sup>?
- Have you finished preparing [the stuff you're talking about tomorrow]?

## 本発表の仮説

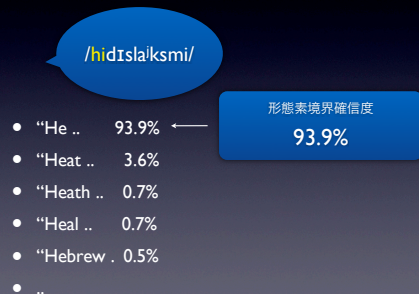
- 接頭辞は（平均的には）処理にくい
- しかし、実在する接頭辞には、処理の難しさを軽減するような性質が備わっている
- 本発表ではこれを定量的に調べる

## 本発表の手法： 理解モデルの定量化

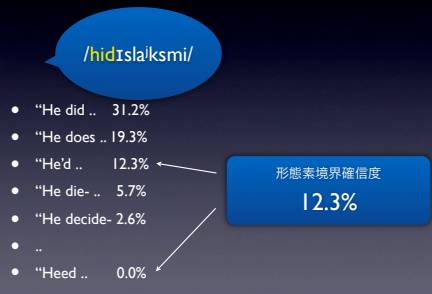
## 本発表の手法



## 本発表の手法



## 本発表の手法



## 本発表の手法

- 架空の接頭辞型言語（接頭辞の処理が困難と予想）
- 逆再生英語
- 実在する接頭辞型言語（接頭辞の困難はないと予想）
- ワルマン語（トリチェリ語族）
- スワヒリ語（ニジェール=コンゴ語族）

## 本発表の手法

- 各言語につき 1,000 文を生成
- 1,000 文を理解プログラムに音素列として与え、形態素列として復元させる
- 形態素境界の確信度を調べ、直前の形態素の種類（接頭辞か否か）ごとに集計

## 逆再生英語

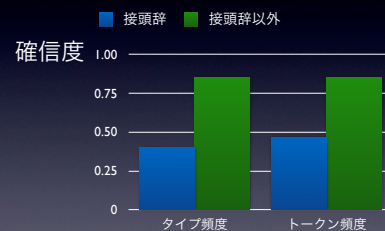
## 逆再生英語

- 音素単位で英語の並び順を逆転したもの
  - im s-kal ih.  
me 3SG.PRES-like he  
'He likes me.'
- Google Web N-gram (Brants and Franz 2006) と CELEX2 (Baayen et al. 1996) からデータを構築

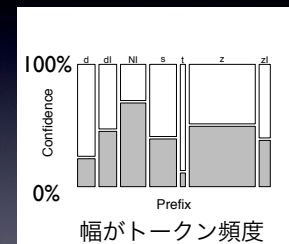
## 逆再生英語

- 英語は接尾辞型言語なので、逆再生英語は**接頭辞型言語**である
- 逆再生英語は実在する言語ではない。従って、**接頭辞の処理の難しさを軽減**するような性質は備わっておらず、**理解が困難**であることを予想

## 形態素境界 平均確信度 (逆再生英語)



## 形態素境界 平均確信度 (逆再生英語)

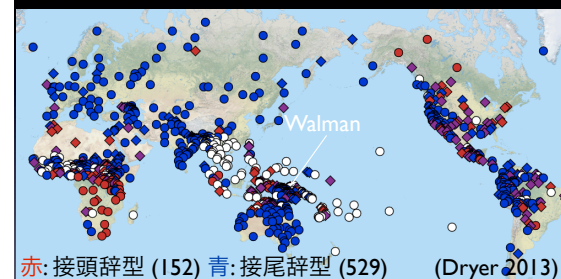


## 逆再生英語

- 逆再生英語の接頭辞はなぜ難しいのか
  - 主要な接頭辞 s-, d- は、たまたまその音素から始まる語が多い
  - 異形態 z-, zI- および d-, dI- の存在
  - これらの接頭辞はゼロと交替する

## ワルマン語

## ワルマン語



# ワルマン語

- パプアニューギニア (トリチェリ語族)
- WALS では weakly prefixing
- 約 25,000 語から成る談話資料を使用 (Matthew Dryer 氏の好意による)
- 本発表の文法記述は Brown and Dryer (2008) による

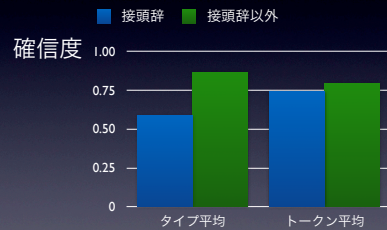
# ワルマン語

- 主語-(命令)-(目的語)-語幹-(適用)-(目的語)
- *Nyanam y-ete-y, ..*  
children 3PL-see-3PL, ..  
Children saw them

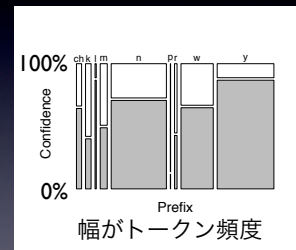
# ワルマン語

主語	単数	複数	目的語	単数	複数
1	m-	k-	1	p-	p-
2	n-	ch-	2	ch-	ch-
3	n- (m.) w- (f.) l- (dimin.)	y-	3	r- (refl.) -n (m.) ∅ (f.) -l (dimin.)	-y

## 形態素境界 平均確信度 (ワルマン語)



## 形態素境界 平均確信度 (ワルマン語)

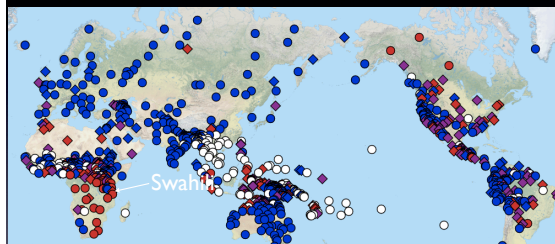


# ワルマン語

- ワルマン語の接頭辞はなぜそれほど難しくもないか
  - 動詞には常に接頭辞が現れ、しかも接頭辞の長さは常に1音素である
  - 接頭辞によってまれな音素連続を生じる。
    - *Mnon n-m-p-klwara.*  
GEN.3SG 3SG.M-IMPER-1.OBJ-deceive  
'Let him lie to me.'
  - SVO 語順のため、主語が明示されたときは動詞接頭辞が予測できる

# スワヒリ語

## スワヒリ語



赤: 接頭辞型 (152) 青: 接尾辞型 (529) (Dryer 2013)

# スワヒリ語

- WALS では weakly prefixing
- 接頭辞型言語のうち、おそらく最大のコーパスをもつ言語
- データに Helsinki Corpus of Swahili の書籍部分 (約100万語) を利用

## スワヒリ語

- 主語-時制-(関係節)-(目的語)-語幹-(拡張接尾辞)-終母音

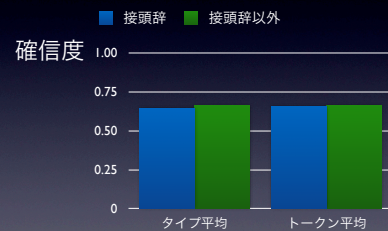
• *Kitabu ni-li-cho-m-som-esh-a mtoto*

book 1/2.SG1.SBJ-PAST-7/8.SG.REL-1/2.SG3.OBJ-read-CAUS-IND child

'The book which I made a child to read'

(佐藤 2012)

## 形態素境界 平均確信度 (スワヒリ語)



## スワヒリ語

- スワヒリ語の接頭辞はなぜそれほど難しくないのであるか
  - 音節構造が単純であり、基本的に1音節ごとに1つの接頭辞である
  - 一致によるものはある程度予測できる?

## まとめ

- 実在する接頭辞は、それほど処理が困難にならない
- 接頭辞と接尾辞にこのような統計的違いがあるとすれば、接頭辞型言語の存在は「接頭辞は処理が困難」という説の弱点ではなく、傍証となる可能性がある

## まとめ

ありがとうございました

ご質問などは [asaokitan@gmail.com](mailto:asaokitan@gmail.com) へ

## 逆再生英語

接頭辞	トークン頻度	平均確信度
d-	83	.233
di-	84	.461
ɲi-	120	.698
s-	127	.398
t-	24	.114
z-	313	.504
zi-	53	.386
接頭辞	804	
タイプ平均		<b>.399</b>
トークン平均		<b>.464</b>
接頭辞以外	6,864	
タイプ平均		<b>.850</b>
トークン平均		<b>.849</b>

## ワルマン語

接頭辞	トークン頻度	平均確信度
ch-	41	.655
k-	44	.408
l-	10	.885
m-	56	.498
n-	440	.721
p-	2	.119
r-	20	.436
w-	257	.661
y-	452	.886
接頭辞	1,322	
タイプ平均		<b>.585</b>
トークン平均		<b>.740</b>
接頭辞以外	2,989	
タイプ平均		<b>.862</b>
トークン平均		<b>.793</b>

## スワヒリ語 (主語人称接辞のみ詳細を表示)

クラス	動詞接頭辞	トークン頻度	平均確信度
1 (subj)	a-	89	.687
1 (obj)	m-	146	.345
2	wa-	151	.674
3	u-	146	.727
4	i-	62	.551
5	li-	165	.805
6	ya-	42	.724
7	ki-	81	.603
8	vi-	35	.643
9	i-	62	.551
10	zi-	28	.541
11	u-	146	.727
15	ku-	230	.765
16	pa-	3	.406
17	ku-	230	.765
18	mu-	1	.212
接頭辞 (上記以外も含む)		1,632	
タイプ平均			<b>.641</b>
トークン平均			<b>.657</b>
接頭辞以外		5,240	
タイプ平均			<b>.660</b>
トークン平均			<b>.663</b>