Parsing with prosody: towards a computational model of prosodically-informed syntactic parsing in Samoan

Kristine M. Yu
krisyu@linguist.umass.edu

Department of Linguistics, UMass Amherst

Brown LingLangLunch Series
April 19, 2017
Collaborators

Matthew Frelinger  UMass Amherst
Deniz Özyildiz  UMass Amherst
Ed Stabler  UCLA/Nuance
Bringing prosody into the grammar to inform the parser

Why doesn’t Siri really take advantage of prosodic information?:

- Many theories of syntax-prosody interface with many moving parts
- Many interacting, conditioning factors on prosody that may obscure informativity of prosodic information for syntax
Strategy:

- Define and compare computational models which capture fundamental properties that distinguish proposed theories from one another
- Consider how the grammar can be used in a performance model (a parser)
- Start with case studies where syntax is clearly the primary determining factor for prosody
  - Today: a first case study on Samoan
A fine-grained model of the syntax-prosody interface

Indirect and direct reference theories agree that phonological rules refer to cross-categorical relationships rather than specific syntactic categories. (Kaisse and Zwicky 1987, p. 7).

Here for Samoan interface: We place tones by individual rules that refer to specific morphosyntactic constructions.

In Samoan, prosodic constituents do not appear to be unified by sharing some common syntactic relation.
Other fine-grained models of interface

Similar examples from languages where word-level tone overlays apply in specific morphosyntactic contexts: Dogon languages of Mali (Heath and McPherson 2013, McPherson and Heath 2016), also Papuan languages Awa and Usarafu, and Nigerian languages Nkoroo and Kalabari. In Tommo So (Dogon):

- *gamma* ‘cat’ HH; also in ‘three cats’ and ‘the cat’
- but ‘black cat’, ‘one cat’, ‘Sana’s cat’: LL
Ergative-absolutive case-marking patterns

**Definition: ergative case marking**

Case of intransitive subject same as case of transitive object

1. **Case-marking in English**
   1. **Intransitive**: He slept. (*Him slept).  
   2. **Transitive**: He hit him.
Ergative-absolutive case-marking patterns

Definition: ergative case marking
Case of intransitive subject same as case of transitive object

1. Case-marking in English
   1. Intransitive: He slept. (*Him slept).
   2. Transitive: He hit him.

2. Case-marking in Samoan
   1. Intransitive: Him slept. (*He slept).
   2. Transitive: He hit him.
Samoan’s ergative-absolutive case system

- Subject of transitive clause receives “ergative” case, marked with prepositional element [e]
- Subject of intransitive clause and object of transitive clause receive “absolutive” case, said to be unmarked
- Prepositional element [i] marks “oblique” case on indirect objects, locatives, etc.
Basic sentences

(1) **Transitive clause**

na lalaña *(e) le malini le mamanu.
PAST weave **ERG DET marine DET design**

‘The marine wove the design.’
Basic sentences

(1) **Transitive clause**

\[ \text{na } \text{lala}\text{ŋa} *(e) \text{ le malini le mamanu}. \]

PAST weave **ERG DET marine DET design**

‘The marine wove the design.’

(2) **Intransitive clause**

\[ \text{na } \text{ŋalue le malini (i le mamanu)}. \]

PAST work **DET marine OBL DET design**

‘The marine worked (on the design).’
Word order

- Primarily VSO word order in transitive clauses, although VOS possible
- Fronted arguments, with SVO or OVS word orders in clefted structures
Word order

- Primarily VSO word order in transitive clauses, although VOS possible
- Fronted arguments, with SVO or OVS word orders in clefted structures

(3) Fronted argument

[ʔo le malini] na lalaŋa le manamu.
[TOPIC DET marine] PAST weave DET design

‘It was the marine that wove the design.’
Basic word prosody

- Described as a **non-tonal language in a non-tonal language family**

- Basic primary stress pattern: moraic trochee at right edge (Zuraw, Yu, and Orfitelli 2014)

- Primary stress on final vowel if long, otherwise on penultimate vowel
  - (‘manu) ‘bird’ ♦, ma(‘lini)
  - la(‘vaː) ‘energized’ ♦
Rising pitch accent associated to primary stressed syllable ("LH*") (Orfitelli and Yu 2009, Calhoun 2015) appears on every content word, presence seems insensitive to discourse conditions.

Sentence-medial high edge tones ("H-") including before an absolutive (Yu 2011, 2016)
Fieldwork/data

- All consultants from (Western) Samoa
- Primary consultant: born in Samoa, moved to United States at age 15, age 19-23 during elicitations
  - Data collected in 2008-2009, and later in additional elicitation sessions through 2016
Fieldwork/data

- All consultants from (Western) Samoa
- Primary consultant: born in Samoa, moved to United States at age 15, age 19-23 during elicitations
  - Data collected in 2008-2009, and later in additional elicitation sessions through 2016
- Work with 5 additional speakers in trips to Apia, Samoa in November 2011 and to Carson, CA in January 2012
- Work with 4 additional speakers in trips to Auckland, New Zealand in July 2015

Data primarily elicited in *tautala lelei* (formal language) Data primarily elicited out-of-the-blue, unless pronominal elements present, in which case referent introduced
Fieldwork/data

- All consultants from (Western) Samoa
- Primary consultant: born in Samoa, moved to United States at age 15, age 19-23 during elicitations
  - Data collected in 2008-2009, and later in additional elicitation sessions through 2016
- Work with 5 additional speakers in trips to Apia, Samoa in November 2011 and to Carson, CA in January 2012
- Work with 4 additional speakers in trips to Auckland, New Zealand in July 2015
- Data primarily elicited in *tautala lelei* (formal language)
- Data primarily elicited out-of-the-blue, unless pronominal elements present, in which case referent introduced
Fundamental frequency contour: transitive clause

<table>
<thead>
<tr>
<th>Time (s)</th>
<th>0</th>
<th>0.5</th>
<th>1</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fundamental frequency (Hz)</td>
<td>100</td>
<td>125</td>
<td>150</td>
<td>175</td>
</tr>
</tbody>
</table>

- LH* LH*H-
- LH*
- L-L%
- LH* LH*
- H-
- LH*
- L-L%

F0 contour for transitive clause in (14)

Kristine M. Yu www.krisyu.org
F0 contour: intransitive clause

F0 contour for intransitive clause in (2)

Kristine M. Yu  www.krisyu.org  Parsing with prosody in Samoan  15
Where H- tones invariably occur

- **Absolutives**: at right edge of word *preceding* absolutive (Yu 2011, 2016)
Where H- tones invariably occur

- **Absolutives**: at right edge of word *preceding* absolutive (Yu 2011, 2016)

- **Coordination**: at right edge of first coordinate (Orfitelli and Yu 2009, Yu 2011, 2016)
Where H- tones invariably occur

- **Absolutives**: at right edge of word *preceding* absolutive (Yu 2011, 2016)

- **Coordination**: at right edge of first coordinate (Orfitelli and Yu 2009, Yu 2011, 2016)

- **Clefts**: at right edge of fronted argument (Orfitelli and Yu 2009, Yu 2011, 2016; Calhoun 2015)
Where $H$- tones invariably occur

- **Absolutives**: at right edge of word *preceding* absolutive (Yu 2011, 2016)

- **Coordination**: at right edge of first coordinate (Orfitelli and Yu 2009, Yu 2011, 2016)

- **Clefts**: at right edge of fronted argument (Orfitelli and Yu 2009, Yu 2011, 2016; Calhoun 2015)

None of these syntactic environments can coincide: an $H$- can be uniquely traced back to its syntactic source.
Where H- tones invariably occur

(4) **Absolutive**

na lalaŋa *(e) le malini [H- le manamu. PAST weave ERG DET marine ABS DET design

‘The marine wove the design.’
Where H- tones invariably occur

(4) **Absolutive**

\[
\text{na lala}'a *(e) le malini } \textbf{H-} \text{ le mamanu.} \\
\text{PAST weave } \textbf{ERG DET marine } \textbf{ABS DET design}
\]

‘The marine wove the design.’

(5) **Coordination**

\[
\text{na lala}'a *[e) le malini } \textbf{H-} \text{ ma Malu}\] \textbf{H-} \\
\text{PAST weave } [\textbf{ERG DET marine} \textbf{COORD CONJ Malu}] \textbf{ABS} \\
\text{le } \text{ mamanu.} \\
\text{DET design}
\]

‘The marine and Malu worked on the design.’
Where H- tones invariably occur

(4) **Absolutive**

\[
\text{na lalaŋa } *(e) \text{ le malini } \underline{\text{H-}} \text{ le mamanu.} \\
\text{PAST weave } \underline{\text{ERG DET marine}} \underline{\text{ ABS DET design}}
\]

‘The marine wove the design.’

(5) **Coordination**

\[
\text{na lalaŋa } *(e) \text{ le malini } \underline{\text{H-}} \text{ ma Malu]} \underline{\text{H-}} \\
\text{PAST weave } [\underline{\text{ERG DET marine}} \underline{\text{ COORD CONJ Malu]} \underline{\text{ ABS DET design}}
\]

‘The marine and Malu worked on the design.’

(6) **Fronted argument**

\[
[?o \text{ le malini]} \underline{\text{H-}} \text{ na ŋalue i le mamanu.} \\
[\underline{\text{TOPIC DET marine}} \underline{\text{ FRONT}} \text{ PAST work OBL DET design}}
\]
A closer look at the phonetic realization of H-

[Diagram showing differences in fundamental frequency between LH* not followed by H- and LH* followed by H-]

- [LH* not followed by H-] [Diagram showing a flat fundamental frequency line from ma to li to ni]
- [LH* followed by H-] [Diagram showing a rise in fundamental frequency from ma to li to ni with H-]

Kristine M. Yu www.krisyu.org
Homophonous H-’s

- No positive evidence that H-’s are realized differently depending on syntactic structure
- Lengthening at right edges with H-’s compared to baseline when no H- is present
Homophonous H-’s

- No positive evidence that H-’s are realized differently depending on syntactic structure
- Lengthening at right edges with H-’s compared to baseline when no H- is present
- Other examples of “homophonous” tones: overlapping tonal inventory for lexical and grammatical tones in Bantu languages; high tones can be edge tones or pitch accents in head/edge-prominence languages, e.g. Turkish, French
(7)   Sensitivity of prosodic phrasing to speech rate in Calcutta Bengali (Hayes and Lahiri 1991)

a.  (ওমর) (ঁদাদর) (তারা-কে) (দিইেচে) *deliberate speech*
    Armor scarf  Tara-obj gave
    ‘Armor gave a scarf to Tara’

b.  (ওমর ঁদাদর) (তারা-কে) (দিইেচে) *faster speech*

c.  (ওমর) (ঁদাদর তারা-কে) (দিইেচে) *faster speech*

b.  (ওমর ঁদাদর তারা-কে) (দিইেচে) *very rapid speech*
Sample minimal pair from speech rate stimuli

(8) A sample minimal pair from speech rate data set

a. Transitive clause

na laŋona e malini H-le liona i
PAST hear ERG marine ABS DET.SG lion OBL
le aoauli.
DET.SG afternoon

‘The marines heard the lion in the afternoon.’

b. Intransitive clause

na manoŋi H-malini i le liona i le
PAST smelly ABS marine OBL DET.SG lion OBL DET.SG
aoauli.
afternoon

‘The marines were smelly to the lion in the afternoon.’
Invariability in presence of H-: speech rate

(a) subject under slow speech rate

(b) subject under normal speech rate
Invariability in presence of H-: speech rate

(c) subject under normal speech rate  (d) subject under fast speech rate
Properties of Samoan registers

- *tautala lelei* ‘good language’: used in literary contexts and and Westernized contexts, with foreigners
- *tautala leaga* ‘bad language’: used in traditional ceremonies and meetings, with family and friends
- Case markers often dropped (Mosel and Hovdhaugen 1992, Mayer 2001)

(9) Mergers from *tautala lelei* to *tautala leaga*

a. /t/ and /k/ $\rightarrow$ /k/

b. /n/ and /ŋ/ $\rightarrow$ /ŋ/
(10) Transitive sentence minimal pair in *tautala leaga*

a. VSO word order

\[ ηa \ lanøŋa \∅ \ le \ liøŋa \textbf{H-} \ le \ malie. \]

\[
\text{PAST}\, \text{hear}\, \text{ERG DET.SG lion}\, \text{ABS DET.SG shark}
\]

‘The lion heard the shark.’

b. VOS word order

\[ ηa \ lanøŋa \textbf{H-} \ le \ liøŋa \∅ \ le \ malie. \]

\[
\text{PAST}\, \text{hear}\, \text{ABS DET.SG lion}\, \text{ERG DET.SG shark}
\]

‘The lion was heard by the shark.’
Invariability in presence of H-: register

(a) f0 over 3-syll verbs

(b) f0 over 3-syll ARG 1
Other invariability in presence/placement of H-’s

- Word order
- Prosodic length (number of syllables/words in arguments)
- Information structural manipulations (Contrastive and informational focus)
Defining the syntax/prosody interface

- Define beginnings of a formal Minimalist Grammar for Samoan (Yu & Stabler to appear)
Defining the syntax/prosody interface

- Define beginnings of a formal Minimalist Grammar for Samoan (Yu & Stabler to appear)
- Adopt a particular syntactic perspective on Samoan case, but other syntactic alternatives could be defined as well
Define beginnings of a formal Minimalist Grammar for Samoan (Yu & Stabler to appear)

Adopt a particular syntactic perspective on Samoan case, but other syntactic alternatives could be defined as well

Why Minimalist Grammars? Regular derivation trees makes it easy to interface with:
1. Regular (tree) transducer to do spellout
2. Regular transducers defined for the prosodic/phonological grammar
What is the absolutive $H$-?


- **Absolutive** is default, syncretic marking of nom and acc
What is the absolutive H-?


- **Absolutive** is default, syncretic marking of nom and acc
  - Collins: S and P behave different in nominalized clauses: S must be genitively marked, while P can have same marking as in finite clauses
What is the absolutive H-?


- **Absolutive** is default, syncretic marking of nom and acc
  - Collins: S and P behave different in nominalized clauses: S must be genitively marked, while P can have same marking as in finite clauses
  - In nominalizations, H- possible for transitive object P but not for intransitive subject S
What is the absolutive H-?


- **Absolutive** is default, syncretic marking of nom and acc
  - Collins: S and P behave different in nominalized clauses: S must be genitively marked, while P can have same marking as in finite clauses
  - In nominalizations, H- possible for transitive object P but not for intransitive subject S

- Case markers adjoined to arguments in spellout, additional spellout rules for coordination and clefted structures
What is the absolutive H-?


- **Absolutive** is default, syncretic marking of nom and acc
  - Collins: S and P behave different in nominalized clauses: S must be genitively marked, while P can have same marking as in finite clauses
  - In nominalizations, H- possible for transitive object P but not for intransitive subject S

- Case markers adjoined to arguments in spellout, additional spellout rules for coordination and clefted structures

- Verb-initial ordering derived by fronting the VP to a function head F below T after the arguments have been raised out of it (Collins)

- Head movement moves T *na to C* (Collins)
What is the absolutive H-?


- **Absolutive** is default, syncretic marking of nom and acc
  - Collins: S and P behave different in nominalized clauses: S must be genitively marked, while P can have same marking as in finite clauses
  - In nominalizations, H- possible for transitive object P but not for intransitive subject S

- Case markers adjoined to arguments in spellout, additional spellout rules for coordination and clefted structures

- Verb-initial ordering derived by fronting the VP to a function head F below T after the arguments have been raised out of it (Collins)

- Head movement moves T *na* to C (Collins)

- Other ideas about *abs*: (1) material preceding *abs* fronted into high Spec position, (2) *abs* arguments extrapose
Derived syntactic trees (Yu & Stabler to appear)

Absolutive

Coordination

Clefting
Absolutive H- is distinct from other H-’s

- Frequency of usage of absolutive ia currently seems very sporadic, but speakers still have systematic intuitions about distribution
Absolutive H- is distinct from other H-’s

- Frequency of usage of absolutive ia currently seems very sporadic, but speakers still have systematic intuitions about distribution
- ia illicit in syntactic environments where other segmental case markers also illicit (Yu and Özyıldız 2016)
- ia licit only where absolutive H- occurs, not where cleft or coordination H-’s occur (Yu & Özyıldız 2016)
Absolutive illicit in naʔo constructions

Calhoun (2014, 2017) noted that no absolutive H- occurs after naʔo. My consultants don’t accept ia with naʔo either:

(11) **No absolutive ia, H-**

na laŋona e Melina *H- *ia naʔo *H- *ia
PAST hear erg Melina *abs *abs only *abs *abs
le liona.
DET.SG lion

‘Melina heard only the lion.’
Naʔo ‘only’ cannot occur with any case morpheme

(12) **No ergative e**

na laʔona *e naʔo Melina *e H- (ia) le
PAST hear *ERG only Melina *ERG ABS (ABS) DET.SG
liona.
lion

‘Only Melina heard the lion.’

(13) **No oblique i**

na leaʔa H- (ia) Melina *i naʔo *i le
PAST bad ABS (ABS) Melina *OBL only *OBL DET.SG
liona.
lion

‘Melina was bad to only the lion.’
A phonological grammar fragment

- Task: assign stress and prosodically parse sequence of morphemes marked for syllable weight
- Constraint-based grammar to assign stress based on phonological constituency from Zuraw, Orfitelli and Yu (2014)
Parsing a transitive clause in phonological constituents

(14) **Transitive clause**

na lalaŋa *(e) le malini le mamanu.  
PAST weave **ERG DET marine DET design**

‘The marine wove the design.’

Kristine M. Yu www.krisyu.org
Why xfst (Beesley and Karttunen 2003)?

Finite state sequence transducers expressive enough to define diverse phonological formalisms

Easy to interface with the defined minimalist grammars
Hypothesis for origin of absolutive H-

Hypothesis: **Absolutive H-** emerged diachronically from segmental [ia] particle: segmental deletion and reassociation of orphaned tone

- ia is bimoraic and receives penult stress; pitch accent provides source tone for reassociation

```
  H   H
     /\  \
    /   \ 
manu  ia
```
Disfluencies and phonological constituency

- Not clear whether we would want to say H-’s are placed at the right or left edge of a phonological constituent.
- Placement of **abs** H- to the left: could be explained if Samoan tends to group case markers leftward, prosodically.
Disfluencies and phonological constituency

- Not clear whether we would want to say H-’s are placed at the right or left edge of a phonological constituent.
- Placement of **abs** H- to the left: could be explained if Samoan tends to group case markers leftward, prosodically.
- Evidence for this: asymmetric chunking in disfluencies—higher frequency of function words being “enclitic” rather than “proclitic” in disfluencies/hesitations (Himmelman 2014)
fāgogo from The Archive of Māori and Pacific Music

fāgogo. fables from Samoa

- ‘O Sina ma Le ‘Ulafala (Sina and Le ‘Ulafala)
- ‘O Gatalaiuluē (Gatalaiulue)
- ‘O le gata (The snake)
- ‘O le fai (The stingray)
- ‘O Sināmoe‘ilepapa (Sināmoe‘ilepapa)
- ‘O Fila ma Fala (Fila and Fala)
- ‘O Tigilau ma le lupe (Tigilau and the pigeon)
- ‘O Tuivalea ma Tuiatamai (Tuivalea and Tuiatamai)
- ‘O Sinaeusu’imanu (Sinaeusu’imanu)
- ‘O Tulelepuia’ilama (Tulelepuia’ilama)
Mapping between syntactic and prosodic trees

Syntax

Prosody
Mapping between syntactic and prosodic trees

Syntax

Prosody
In Samoan, placement of sentence-medial high edge tones entirely predictable with knowledge of syntactic structure
In Samoan, placement of sentence-medial high edge tones entirely predictable with knowledge of syntactic structure

Opportunity to carefully evaluate assumptions in theories of syntax-phonology interface
In Samoan, placement of sentence-medial high edge tones entirely predictable with knowledge of syntactic structure

Opportunity to carefully evaluate assumptions in theories of syntax-phonology interface

Starting point for defining and comparing computational models of the interface which capture fundamental properties that distinguish proposed theories from one another
Acknowledgements

- **Primary consultants:** John Fruean, Kare’l Lokeni
- **Coordination of Samoan fieldwork:** Gladys Fuimaono, Peone Fuimaono; Coordination of Auckland fieldwork: Jason Brown
- **Adam Albright, Ryan Bennett, Rajesh Bhatt, Seth Cable, Sandy Chung, James Collins, Lyn Frazier, Alice Harris, Larry Hyman, Sun-Ah Jun, Diana Massam, Jim McCloskey, Masha Polinsky, Russ Schuh, Lisa Selkirk, Donca Steriade, Ellen Woolford, Kie Zuraw
- **Funding:** from UCLA Academic Senate’s Committee on Research, Department of Linguistics at University of Maryland College Park and University of Massachusetts Amherst