Information Structure in Spoken Japanese: Particles, Word Order, and Intonation

Natsuko Nakagawa

2016/2/17

Abstract

This thesis investigates the associations between information structure and linguistic forms in spoken Japanese mainly by analyzing spoken corpora. It proposes multi-dimensional annotation and analysis procedures of spoken corpora and explores the relationships between information structure and particles, word order, and intonation.

Particles, word order, and intonation in spoken Japanese have been investigated separately in different frameworks and different subfields in the literature; there was no unified theory to account for the whole phenomena. This thesis investigated the phenomena as a whole in a consistent way by annotating all target expressions in the same criteria and by employing the same analytical framework. Chapter 1 outlines the questions to be investigated and introduces the methodology of this thesis. Chapter 2 reviews the literature of Japanese linguistics as well as the literature on information structure in different languages. Chapter 3 proposes the analytical framework of the thesis. Major findings are discussed in Chapter 4, 5, and 6.

Chapter 4 analyzes the distributions of topic and case particles. It is made clear that so-called topic particles (*wa*, zero particles, *toiuno-wa*, and *kedo/ga* preceded by copula) are mainly sensitive to activation status, whereas case paticles (*ga*, *o*, and zero particles) are sensitive to both focushood and argument structure. While the distinction between *wa* and *ga* gather much attention in traditional Japanese linguistics, the distribution of different kinds of topic and case particles, including zero particles, are analyzed in this thesis.

Chapter 5 studies word order: i.e., clause-initial, pre-predicate, and post-predicate noun phrases. Topical NPs appear either clause-initially or post-predicateively, while focal NPs appear pre-predicatively. Clause-initial and post-predicate NPs are different mainly in activation statuses. The previous literature investigated clause-initial, pre-predicate, and post-predicate constructions in different frameworks; however, there was no unified account for word order in Japanese. The thesis outlines word order in spoken Japanese in a unified framework.

Chapter 6 investigates intonation. While the previous literature mainly concentrates on contrastive focus, this thesis discusses in terms of both topic and focus. It turns out that intonation as a unit of processing and argues that information structure influences on the form of intonation units.

Chapter 7 discusses theoretical implications of these findings. Finally, Chapter 8 summarizes the thesis and points out some remaining issues and possible future studies.

Contents

1 Introduction

Goal of the study

- Investigate relations btw **information structure** (IS) and **linguistic forms**

- Propose a cross-linguistic method of corpus investigation

(1) A1: tanosii-ne: **ongaku** (← post-predicative, zero-coded)
         fun-FP   music
         'It's a lot of fun, music.'
     B2: un tanosii-yo:
         yes fun-FP
         'Yeah, (it's) fun.'
     A3: ii-na: tyotto ongaku-bu-ni   hairi-takat-ta-na:
         good-FP a.bit   music-club-DAT enter-want-PAST-FP
     C4: ima-kara-de-mo   **gassyoo-dan** doo (← pre-predicative, zero-coded)
         now-from-at-also chorus-club   how
         'How about the chorus club from now?'   (`chiba0332: 72.69-81.30`)

Proposal

- **Multi-dimensional analysis** of:

  - **Particles** (*toiuno-wa, wa, ga, ga/kedo o, & Ø*)

  - **Word order** (clause-initial, pre-predicate, & post-predicate elements)

  - **Intonation** (phrasal vs. clausal IU)

  - in spoken Japanese

  - in terms of IS

What is IS?   "[T]he utterance-internal structural and semantic properties reflecting the relation of an utterance to the discourse context, in terms of the discourse status of its content, the actual and attributed attentional status of the discourse participants, and the participants' prior and changing attitudes (knowledge, beliefs, intentions, expectations, etc.)" (Kruijiff-Korbayová & Steedman, 2003, 250).

Background

- Roots of studies on IS (see Kruijiff-Korbayová & Steedman, 2003)

  - **Formal approach** (Russell, 1905; Strawson, 1950, 1964; Chomsky, 1965; Jackendoff, 1972; Selkirk, 1984; Rooth, 1985; Rizzi, 1997; Erteschik-Shir, 1997, 2007; Büring, 2007; Ishihara, 2011; Krifka & Misan, 2012; Endo, 2014)

  - **Functional approach** (Mathesius, 1928, 1929; Sgall, 1967; Firbas, 1975; Bolinger, 1965; Halliday, 1967; Kuno, 1973; Gundel, 1974; Chafe, 1976, 1994; Prince, 1981; Givón, 1983; Tomlin, 1986; Lambrecht, 1994; Birner & Ward, 1998, 2009)

  - **Both traditions** (Vallduví, 1990; Steedman, 1991; Vallduví & Vilkuna, 1998)

- Roots of studies on IS

- **Japanese linguistics** (Matsushita, 1928; Yamada, 1936; Tokieda, 1950/2005; Mikami, 1953/1972, 1960; Onoe, 1981; Kinsui, 1995; Kikuchi, 1995; Noda, 1996; Masuoka, 2000, 2012)
- **Corpus approach** (Hajičová, Panevová, & Sgall, 2000; Calhoun, Nissim, Steedman, & Brenier, 2005; Götze et al., 2007; Chiarcos et al., 2011)

2 Background

Particles

- Distribution of **zero particles** still not clear enough (Tsutsui, 1984; Matsuda, 1996; Fry, 2001)

- *Ga* & *o* sometimes code focus and need to be discussed in terms of IS

- *Wa* & **other topic particles** need coherent explanation

- This sudy

  - Captures **distributions of zero and overt particles as a whole** in terms of IS

Word order

- Different theories focus on different aspects on word order

- **Generative grammar**: "scrambling", more recently left periphery (Saito, 1985; Endo, 2014)

- **Functional linguistics**: post-predicate construction (Ono & Suzuki, 1992; Fujii, 1995; Ono, 2007)

- This study

  - Provides **coherent theory** to explain the whole phenomena

Intonation

- Most studies concentrate on **focus** (e.g., Kori, 2011)

- Corpus studies on **intonation units** rely on impressionistic approach (Iwasaki, 1993; Matsumoto, 2000; Nakagawa, Yokomori, & Asao, 2010)

- This study

  - Employs **IU of clear definitions**
  - Investigate **both topic and focus**

3 Framework

3.1 Theoretical framework

Correlating features of IS

- Topic & focus are **multi-dimensional**; i.e., **bundles of features**

	topic	focus
a.	presupposed	asserted
b.	active	inactive
c.	definite	indefinite
d.	specific	non-specific
e.	animate	inanimate
f.	agent	patient
g.	inferable	non-inferable

(Givón, 1976; Keenan, 1976; Comrie, 1979, 1983)

Topic

- Definition

  - Topic is a discourse element that the speaker assumes or presupposes to be **shared** (known or taken for granted) and **uncontroversial** in a given sentence both by the speaker and the hearer.

- **Shared**: evoked, inferable, declining, or unused in given-new taxonomy (Prince, 1981)

- **Uncontroversial**: cannot be repeated after *hee* or *aha*; cannot be negated in a normal way

Focus

- Definition

  - Focus is a discourse element that the speaker assume to be **news** to the hearer and possibly **controversial**. S/he wants the hearer to learn the relation of the presupposition to the focus by his/her utterance. In other words, focus is an element that is asserted.

- **News** & **Controversial**: can be repeated after *hee* or *aha*; can be negated in a normal way

3.2 Corpus

Corpus

- *the Corpus of Spontaneous Japanese* (CSJ; Maekawa, 2003; Maekawa, Kikuchi, & Tsukahara, 2004)

Table 1: Corpus used in this study

ID	Gender (age)	Theme	Length (sec)
S00F0014	F (30-34)	Travel to Hawaii	1269
S00F0209	F (25-29)	Being a pianist	619
S00M0199	M (30-34)	Kosovo War	580
S00M0221	M (25-29)	Working at Sarakin	654
S01F0038	F (40-44)	Luck in getting jobs	628
S01F0151	F (30-34)	Trek in Himalayas	765
S01M0182	M (40-44)	Boxing	644
S02M0198	M (20-24)	Dog's death	762
S02M1698	M (65-69)	Dog's death	649
S02F0100	F (20-24)	Rare disease	740
S03F0072	F (35-39)	A year in Iran	816
S05M1236	M (30-34)	Memories in Mobara	832

Table 2: Activation status in the corpus

Activation status	Given-new taxonomy	Corpus annotation
Active	Evoked	Given
Semi-active	Declining	
Semi-active	Inferable	
Inactive	Unused	New
Inactive	Brand-new	

Corpus annotation

- Procedure

(2) a. Identification of **argument structure**, discourse elements, and zero pronoun
     b. Classification of discourse elements: Discourse elements are classified into categories based on what they refer to.
     c. Identification of anaphoric relations: The link between the anaphor and the antecedent is annotated.
     d. **Activation statuses** are calculated automatically based on anaphoric relations.
     e. Other features are examined manually on each occasion.

Corpus annotation

- **Given**, if the element in question has the antecedent

- **New**, otherwise

Table 3: Topic marker vs. activation status

Activation status	Given-New taxonomy	Topic	Focus
Strongly active	Evoked	(Zero pronoun)	–
		(Overt pronoun) *toiuno-wa, wa,* Ø	
Active	Evoked	*toiuno-wa, wa,* Ø	case markers, Ø
Semi-active	Inferable	*wa,* Ø	
Semi-active	Declining	COP-*kedo/ga,* Ø	
Inactive	Unused		
Inactive	Brand-new	–	

4 Particles

Summary Table 3

Results Figure 1 & 2

4.1 Topic

Toiuno-wa

- Active elements with explicit antecedent

  (3) a. syokugyoo-ni taisite-no <u>un</u>-toiu   koto-o tyotto o-hanasi
          job-to   towards-GEN fortune-QUOT thing-*o* a.bit   PLT-talk
          si-tai-to   omoi-masu
          do-want-QUOT think-PLT
          'I would like to talk a bit about fortune in job.'
      b. de **un-toiuno-wa**   maa iroirona un-ga   aru-to
          then fortune-QUOT-*toiuno-wa* FL   various fortune-*ga* exist-QUOT
          omou-n-desu-keredomo
          think-NMLZ-PLT-though
          'I guess there are various kinds of fortunes...'   (S01F0038: 0.53–8.70)

- Active elements with implicit antecedent

  (4) a. ee <u>sekai-taitoru-sen-o-desu-ne</u> ee <u>terebi-de</u> <u>mi-masi-ta</u>
          FL world-title-fight-*o*-PLT-FP   FL TV-by   watch-PLT-PAST
          '(My friend and I) watched a world title match on TV.'
      b. ...
      c. watasi-zisin gu -wa ee amari   koo **supootu-kansen-teiunowa** tyotto
          1SG-self   FRG -*wa* FL not.really FL sport-watching-*toiunowa*   FL
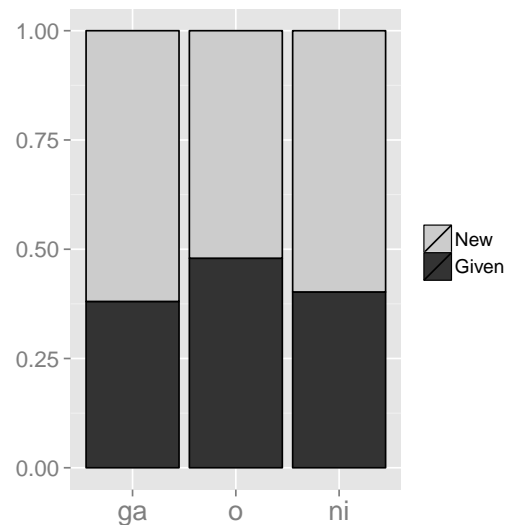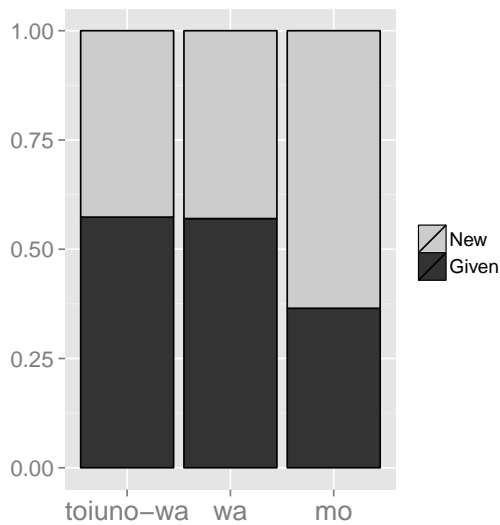   si-nakat-ta-n-desu-ne
          do-NEG-PAST-NMLZ-PLT-FP

1.00

0.75

0.50 — New / Given

0.25

0.00

toiuno–wa   wa   mo

1.00

0.75

0.50 — New / Given

0.25

0.00

ga   o   ni

Figure 1: Topic marker vs. information status (ratio)

Figure 2: Case marker vs. information status (ratio)

'I myself hadn't watched any kinds of sports.'   (S01M0182: 52.77-79.62)

- Semi-active elements (rare)

(5) a. (The speaker moved to Iran when she is a middle school student.)
  b. (The school for Japanese students in Iran was small but she had a lot of fun there.)
  c. eeto iran-no **kikoo-tteiuno-wa** tomokaku kansoo si-tei-masi-te
      FL Iran-GEN *climate-toiuno-wa* at.any.rate dry do-PROG-PLT-and
      'Uh, the climate in Iran was very dry...'   (S03F0072: 178.31-181.65)

- Semi-active elements (*toiuno-wa*-coding unnatural)

(6) a. To start Himalaya trekking, you first fly to a village called Lukla whose elevation is 2600 meters.
  b. From that village, we started trekking.
  c. sono rukura-no mura-nan-desu-ga
      that Lukla-GEN village-NMLZ-PLT-though
      'Regarding that Lukla village,'
  d. **hikoozyoo**-{wa(/??-**toiuno-wa**)} hontooni yama-no naka-ni
      airport-*wa*(/-*toiuno-wa*) really mountain-GEN inside-in
      ari-masi-te
      exist-PLT-and
      'the airport is really in a mountainous area.'   (S01F0151: 179.50-191.39)

Wa

- Active elements

8

(7) a. There is a dish called <u>chelow kebab</u>.
     b. de **sore-wa** eeto gohan-ni eeto bataa-o maze-te
         and that-*wa* FL rice-to FL butter-*o* mix-and
         'That, you mix rice with butter...'
     c. on top of that you put spice,
     d. on top of that you put mutton,
     e. you mix it and eat it.
     f. There were many dishes of this kind.
     g. **sore-wa** kekkoo   sonnani hituzi-no oniku-no kusasa-mo
         that-*wa* to.some.extent not.really sheep-GEN meat-GEN smell-also
         naku-te
         not.exist-and
         'It did not have smell of mutton...'
     h. I thought it was delicious.   (S03F0072: 446.03–471.72)

- Semi-active elements

(8) a. ee toaru ryokoo-sya-ni   ano itioo   <u>nyuusya</u> kimari-masi-ta
         FL certain travel-company-DAT FL tentatively admission decide-PLT-PAST
         'A certain travel company admitted me to work there.'
     b. ...
     c. hizyooni **siken-wa** muzukasikat-ta-to   ima-mo oboe-teori-masu
         very   exam-*wa* difficult-PAST-QUOT now-also remember-PROG-PLT
         '(I) still remember that the exam was very hard.'
         (S01F0038: 231.34–241.96)

- Semi-active elements (accommodated)

(9) a. tada soko-kara saki-wa   ano dono sigoto-mo soo-da-to   omou-n-desu-ga
         but that-from ahead-*wa* FL which job-also   so-COP-QUOT think-NMLZ-PLT

         'But, after the admission, I guess this is the same in all kinds of jobs,'
     b. yume-to   genzitu-tte   iu-n-desu-ka
         dream-and reality-QUOT call-NMLZ-PLT-Q
         'people might call it (the difference between) dream and the reality,'
     c. **gyappu-wa** kanari ari-masi-te
         gap-*wa*   very   exist-PLT-and
         'there was a gap (between what I expected and the reality).'   (S01F0038:
         265.11–270.98)

- *Wa* sometimes "forces" the hearer to accommodate the assumption.

Contrastive *wa*

- Contrastive *wa*-coded elements = semi-active elements

(10) a. de doitu-toiu   kuni-wa   hizyooni ano uu inu-ni   e sumi-yasui
and Germany-QUOT nation-*wa* very   FL FL dog-DAT FL live-easy
kuni-desu
nation-COP.PLT
'Germany is easy for dogs to live in.'

b. tatoeba   aa resutoran-de-mo   anoo **tinomigo-wa**
for.example FL restaurant-at-also FL   infant-*wa*
haire-nai-yoona   resutoran-mo   **inu-wa** haireru-to
enter.can-NEG-such.as restaurant-also dog-*wa* enter.can-QUOT
'For example, restaurants that infants are not allowed to get in, uh, dogs can
get into them.'   (S02M1698: 243.46–256.10)

- (Creatures who can get into) restaurant

  - *inu* 'dog'
  - *tinomigo* 'infant'

copula + *kedo/ga*

- Semi-active declining elements

(11) a. kore-kara ano mokuhyoo-tteiuno-ga ari-masi-te
this-from FL goal-toiuno-*ga*   exist-PLT-and

b. ma sore-wa ookiku wake-te   hutatu aru-n-desu-keredomo
FL that-*wa* roughly divide-and two   exist-NMLZ-COP.PLT-though

c. ma <u>meesee</u>-no bubun-to <u>sigoto</u>-tteiu bubun-ga ari-masi-te
FL fame-GEN part-and job-called   part-*ga*   exist-PLT-and
'I have two goals: one is for <u>fame</u> and the other is for <u>job</u>.'

d. Concerning <u>fame</u>,

e. I have been participating in various piano competitions.

f. So far the best award I received was the fourth best play in the China-Japan
International Competition.

g. Beyond that, I would like to receive higher awards.

h. Titles matters a lot for pianists, so I will work hard.

i. de   ato-wa   **sigoto-no bubun-nan-desu-keredomo**
then remaining-*wa* job-GEN   part-NMLZ-COP.PLT-though
'Concerning the other one, job,'

j. to receive heigher wages...   (S00F0209: 495.77–534.04)

- Active but not established as topic?

(12) a. While we trek on the Everest Trail, the cook made us lunch on the way,

b. ato-wa   <u>thii-taimu-tte</u> it-te
remaining-*wa* tea-time-QUOT call-and
'(we) called (it) tea time,'

c. totyuu-de   tyotto bureeku suru koto-ga aru-n-desu-keredomo
on.the.way-at a.bit break   do   thing-*ga* exist-COP.PLT-though
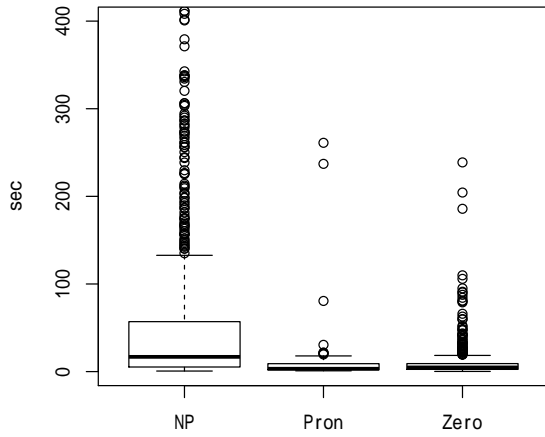'in addition, we had tea time to take a break while we climb the mountain,'

10

Figure 3: Anaphoric distance vs. expression type (all)
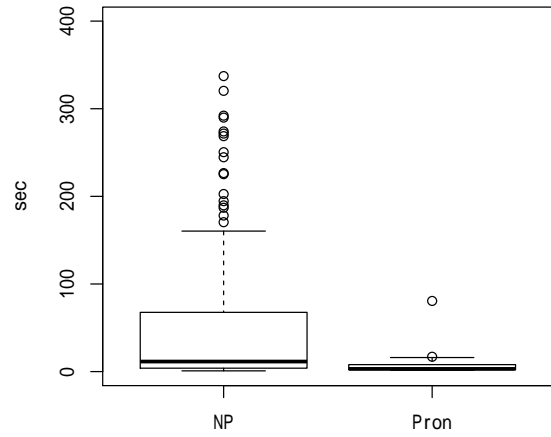
Figure 4: Anaphoric distance vs. expression type (coded by topic markers)

      d.   ee kanari ee souiu tuaa-de ki-teiru-tteiu insyoo-o son'nani atae-nai-de aruku koto-ga deki-masi-ta
FL very.much FL such group.tour-with come-PROG-QUOT impression-*o* so.much give-NEG-and walk thing-*ga* can-PLT-PAST
'we walked without feeling that we were in a big group.'

      e.   de   kono **thii-taimu-nan-desu-keredomo**
and this   tea-time-NMLZ-COP.PLT-though
'And at this tea time,'

      f.   'in this place of high elevation, there is a possibility of altitude sickness, so...'

      g.   'water is very important.'   (S01F0151: 323.00-349.56)

- Inactive unused elements (not attested because of the nature of corpus)

   (13)   Context: Y knows that H, Y's roommate, keeps ice cream in the fridge but saw Taro, another roommate, eat all of H's ice cream after H had left for school. Y wants to tell H this fact when Y sees H in school.

        Y: sooieba   **aisu-{da-kedo/??wa}**   taro-ga tabe-tyat-ta-yo
by.the.way ice.cream-{COP-though/TOP} Taro-*ga* eat-PFV-PAST-FP
'By the way, Taro ate up (your) ice cream.'

Strongly active elements

- Anaphoric distance

    – Distance btw the element in question and the antecedent (sec.)

    – Figure 3 and 4

11

Table 4: Overt vs. zero case markers

	A	S		P
		Agent	Patient	
Non-Contrastive Focus	*ga*	*ga*	*ga*/Ø	Ø
Contrastive Focus or Formal Speech	*ga*	*ga*	*ga*	*o*

Table 5: Word order vs. activation status

Activation status	Given-New taxonomy	Topic	Focus
Strongly active	Evoked	(Zero pronoun) (Overt pronoun) Post-predicate	–
Active	Evoked	Clause-initial	Pre-predicate
Semi-active	Inferable		
Semi-active	Declining		
Inactive	Unused		
Inactive	Brand-new	–	

4.2 Focus

Distribution of case markers   Table 4

4.3 Discussion

Findings

- Different **topic markers** are sensitive to different **activation statuses**

- **Case and zero particles** have **split-intransitive distribution**

5 Word order

Summary   Table 5

Results   Figure 5 & 6

5.1 Topic

Clause-initial elements

- All kinds of topics can appear clause-initially

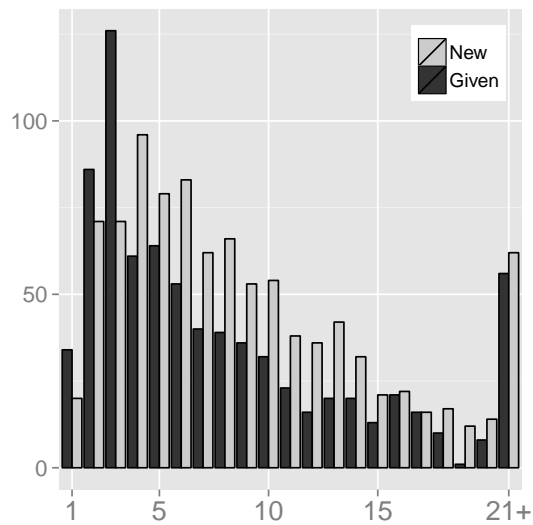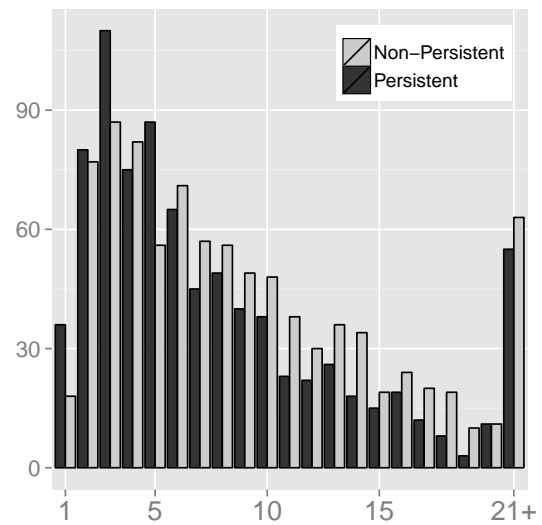- Shared elements appear clause-initially

Figure 5: Word order vs. infoStatus

Figure 6: Word order vs. persistence

(14) a. 'Our grandfather likes sweets.'
      b. yoku pan-ya-san-de   kasi-pan-o   kat-te
          often bread-store-HON-LOC   sweet-bread-*o*   buy-and

          kuru-n-desu-ga
          come-NMLZ-COP.PLT-though
          '(He) often buys sweet bread and comes home,'
      c. e n **sore-o** i maa yoowa oziityan-wa   issyookenmee
          FL FRG that-*o*   FRG FL in.a.word grandfather-*wa* trying.best

          [taberu]-n-desu-keredomo
          eat-NMLZ-COP.PLT-though
          'that, he tries his best to eat it, but'
      d. he cannot eat all and
      e. gives leftovers to the dog...   (S02M0198: 244.48-262.82)

- Unshared elements do not appear clause-initialy

(15) a. desukara daitai   iti-niti-ni   ni-rittoru-no **mizu-o** tot-te   kudasai-to
          so   approximately one-day-for two-litter-GEN water-*o* drink-and please-QUOT
          iw-are-te
          tell-PASS-and
          'So we were told to drink two litters of water per day,'
      b. syokuzi-no toki-wa kanarazu magukappu-de ni-hai-bun-no   **mizu-o**
          meal-GEN time-*wa* surely   mug-with   two-cup-amount-GEN water-*o*
          nomi-masu-si
          drink-PLT-and
          'whenever we have meal, we drink two cups of water,'
      c. totyuu   totyuu-de-mo   kanarazu **mizu-o** ho anoo nomi-taku-naku-temo
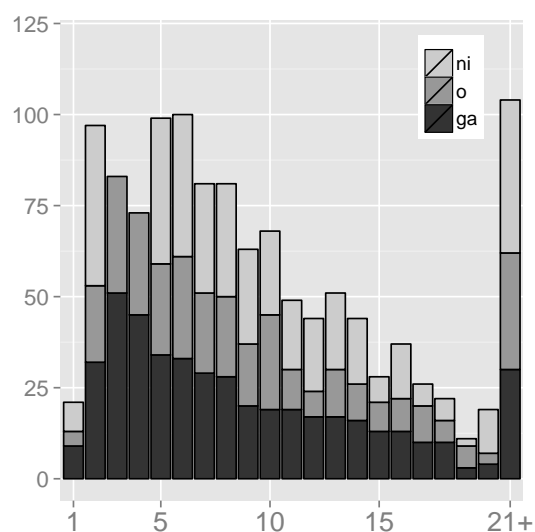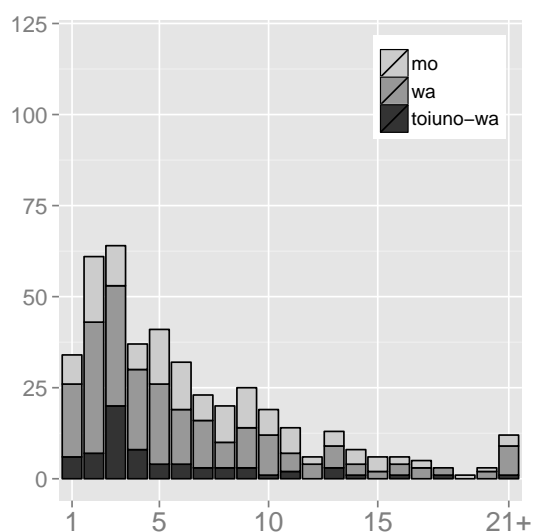   on.the.way on.the.way-LOC-also surely   water-*o* FRG FL   drink-want-NEG-even.if

13

Figure 7: Order of arguments coded by topic markers

Figure 8: Order of arguments coded by case markers

'also on the way, even if we didn't want to drink water,'
d.   'we were forced to drink (water).'
e.   they think that drinking water is very important.   (S01F0151: 339.78–366.29)

(16)   a.   'Also for Kilauea, (we) bought a map and'
b.   de   zibun-tati-de ma rentakaa **kuruma-o** <u>tobasi</u>-te e iki-masi-ta
       then self-PL-by   FL rent-a-car car-*o*   drive-and FL go-PLT-PAST
       '(we) drove there by rent-a-car by ourselves.'
       (83.52 sec talking about the mountain.)
c.   de anoo jibun-no koko koko-de tyotto tome-te miyoo-to omot-ta   toko-ni
       and FL   self-GEN FRG here-LOC a.bit   stop-and try-QUOT think-PAST place-DAT
       koo   **kuruma-o** <u>tome</u>-te
       this.way car-*o*   stop-and
       'At the place (we) wanted to stop, (we) stopped the car,'
d.   you can take pictures and so on.   (S00F0014: 843.23–940.34)

Topic-coded elements appear clause-initially   Figure 7 & 8

Pronouns appear clause-initially   Figure 9 & 10

Strongly active topics appear post-predicatively

(17)   R: nani yat-teru-no   **kono hito**
              what <u>do</u>-PROG-NMLZ this   person
              'What is (he) doing, this person?'   (D02F0028: 193.30–194.45)

(18)   L: sangurasu-toka <u>kake</u>-te-masu-yo-ne **terii itoo-tte**
              sunglasses-HDG wear-PROG-PLT-FP-FP Terry Ito-QUOT
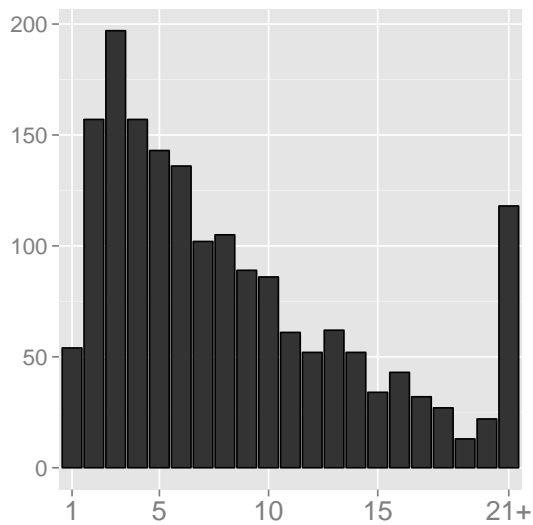   '(He) is wearing sunglasses, isn't he, Terry Ito?'   (D02F0015: 359.17–362.42)

14
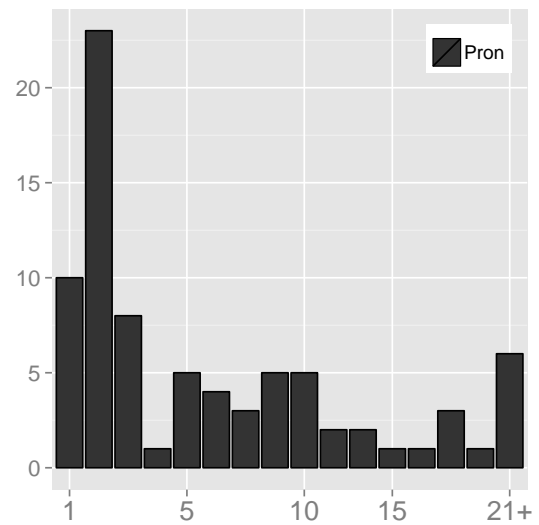
Figure 9: Order of all elements

Figure 10: Order of pronouns

Table 6: RD of post-predicate elements

	Single-contour	Double-contour
RD	**6.9**	39.7

- Mostly appear in **conversations** but not frequently in **monologues**

- Measured referential distance (Givón, 1983) btw the element in question and the antecedent by inter pausal unit

- Post-predicate elements are most frequently pronouns (Nakagawa, Asao, & Nagaya, 2008)

- RD of post-predicate elements is smaller than that of elements before predicate

- Post-predicate elements are "**strongly active**"

5.2   Focus

Pre-predicate elements   Figure 11 & 12

Focus appear pre-predicatively

Table 7: RD of elements before predicate (monologue)

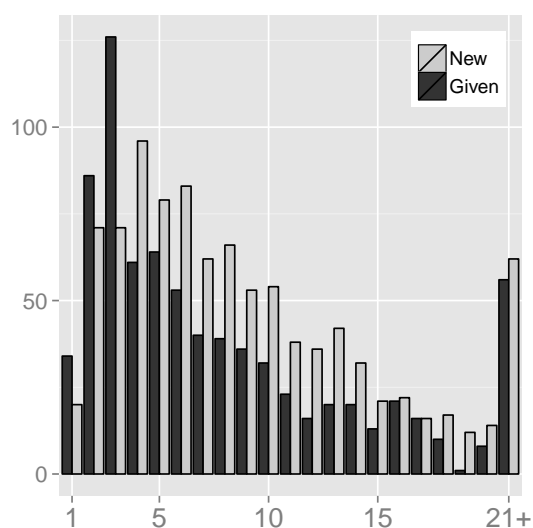|	1	2	3
RD	**20.9**	**23.0**	**41.1**

15

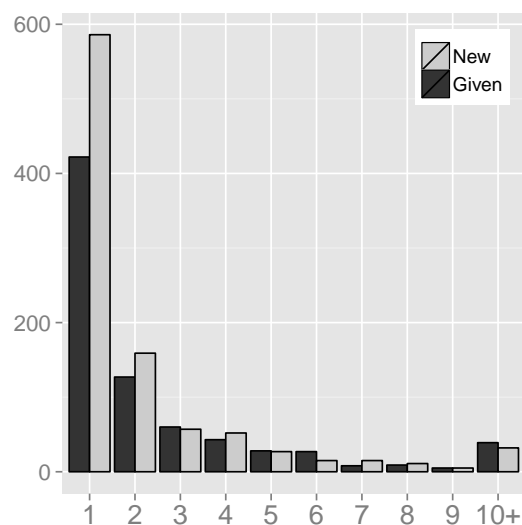Figure 11: Word order vs. information status

Figure 12: Distance from predicate vs. InfoStatus
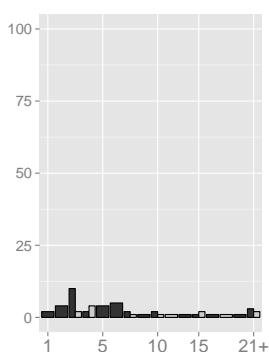

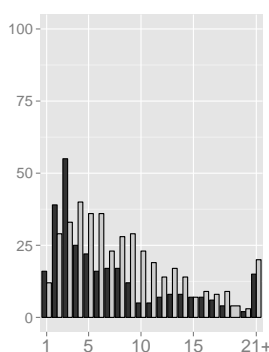Figure 13: Word order of A

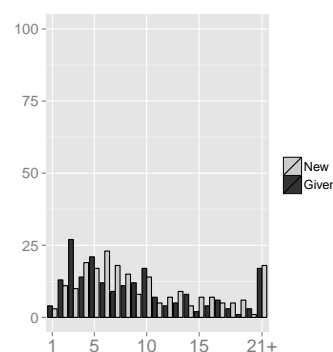Figure 14: Word order of S

Figure 15: Word order of P

(19)   de   ee sono ri-too-no   hoo-ni   sono **kyoomi-o** <u>moti</u> hazime-masi-te
        then FL FL   remote-island-GEN direction-DAT FL interest-*o* have start-PLT-and
        '(We) are started to be interested in remote islands (in Hawaii).'   (S00F0014:
        149.92-153.33)

(20)   a.   sono kontorasuto-toiuno-wa nanka   totemo koo ekizotikku-to-iu-ka
             that contrast-*toiuno-wa*   somehow very   such exotic-QUOT-say-Q
             'The contrast (the color of black and blue) is very exotic, I would say,'
        b.   husigina   **kanzi**-ga   <u>si</u>-masi-te
             mysterious impression-*ga* do-PLT-and
             'the impression was mysterious.'   (S00F0014: 1042.88-1047.03)

- The tendency holds regardless of word order

16

5.3 Discussion

Discussion

- Findings

  (21)   a.   [Clause-init]$_{\textbf{Top}}$ [Pre-predicate Predicate]$_{\textbf{Foc}}$
                  b.   [Pre-predicate Predicate]$_{\textbf{Foc}}$ [Post-predicate]$_{\textbf{Top}}$

    – Confirmed well-known tendency by actual spoken data

- Information-structure continuity principle

  (22)   A unit of IS is continuous in a clause; i.e., elements which belong to the same unit are adjacent with each other.

Discussion

- Clause-initial elements

    – Anchor to the previous discourse (classic observation)
    – Announce the referent of following zero pronouns

- Post-predicate elements

    – Best position for the intonational reason (Mithun, 1995)

- Pre-predicate elements

    – Tied to the predicate

6 Intonation

Phrasal vs. clausal IUs

- Dependent variables

  (23)   a.   **Phrasal IU**: ⬚NP ⬚Predicate
                  b.   **Clausal IU**: ⬚NP Predicate

Summary   Table 8

Results   Figure 16 & 17

Table 8: Intonation vs. activation status

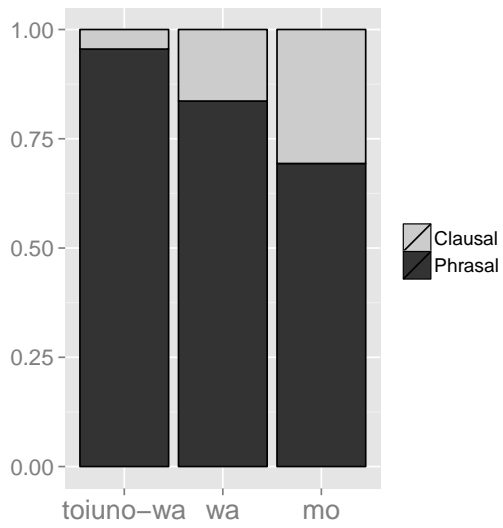| Activation status	Given-New taxonomy	Topic	Focus
Strongly active	Evoked	(Zero pronoun) Clausal IU	–
Active	Evoked	Phrasal IU	Clausal IU
Semi-active	Inferable	Phrasal IU	Clausal IU
Semi-active	Declining		
Inactive	Unused		
Inactive	Brand-new	–	

Figure 16: Intonation unit vs. topic marker
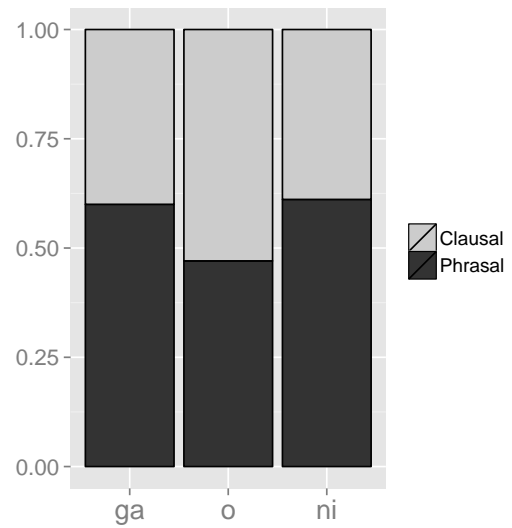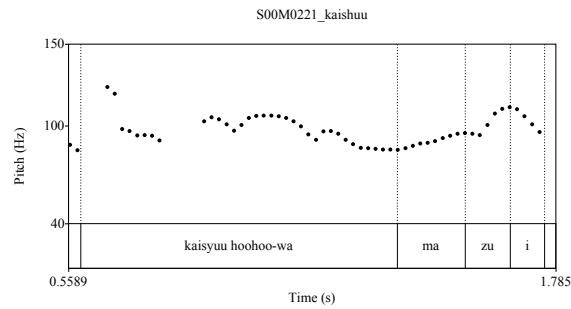

Figure 17: Intonation unit vs. case marker

6.1 Topic

Topics in phrasal IUs

- Pitch reset

  (24)　koo　　　　it-ta　　　‖ **kaisyuu　hoohoo-wa** ‖ mazui-to　　　‖
  　　　this.way say-PAST ‖ collecting method-*wa* ‖ wrong-QUOT ‖
  　　　'This way of collecting (debt) is wrong...'　　　(S00M0221:　580.21–582.06)

- Pitch reset & pause

  (25)　**teema-wa** ‖ hawai-too-no　　　sizen-no　　subarasisa-to　　‖ tabi-no　　‖
  　　　theme-*wa* ‖ Hawaii-island-GEN nature-GEN wonderfulness-and ‖ travel-GEN ‖
  　　　tanosisa-nituite-desu
  　　　pleasure-about-COP.PLT
  　　　'The topic (of this talk) is the wonderful nature and fun travelling in Hawaii
 　　　island.'　　　　　　　　　　　　　　　　　　(S00F0014:　0.30–6.08)
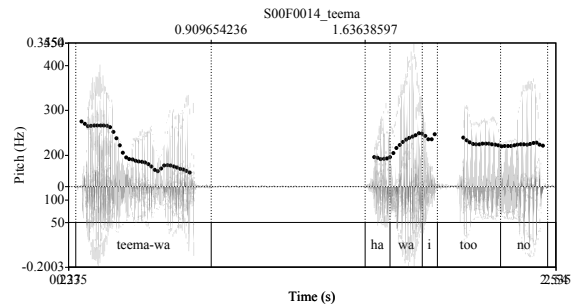
- Exceptions (topics in clausal IUs)
  - Pitch range of topic is larger than the predicate because:
  - Topic is contrasted
  - Clause should form a single unit for other reasons (such as embedded or inserted clause)

Strongly active topics in putative clausal IUs

- No final mora lengthened & no rising

- No pitch reset at the beginning of the following IU

- No pause

- Strongly activated elements without F_0 peak

- Especially pronouns are cliticized.

- Element and predicate form a single processing unit.

(26)   **sore-wa** ‖ nan-daroo-to   omot-te   ‖
        that-*wa* ‖ what-COP.INFR-QUOT think-and ‖
        '(I) was wondering what it was...'   (S00F0014: 654.06–655.18)
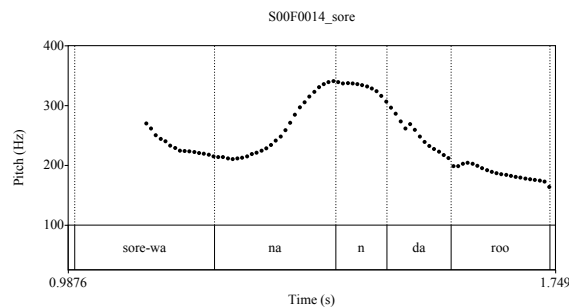
- No time to plan the following utterance

- Magic number is too small (see also Cowan, 2000, 2005)

- Crosslinguistically, unstressed pronouns easily change into clitics, then into affixes. (Givón, 1976)

6.2   Focus

Foci in clausal IUs

- No pitch reset & no pause & no lengthening & no rising

  (27)   a.   our way of collecting debt might be problematic,
          b.   oo mina-san   ‖ zisyuku suru-yooni-to iu ‖ **o-hanasi-ga**
                FL everyone-HON ‖ control do-IMP-QUOT say ‖ PLT-speech-NOM
                <u>de</u>-masi-te   ‖
                come.out-PLT-and ‖
                'somebody proposed that employees should improve the method.' (S00M0221: 503.23–511.02)

S00M0221_ohanashi

- No pitch reset & no pause & no lengthening & no rising

(28) a. ano **puro-raisensu-o**   <u>tori</u>-tai-toka   ‖
          FL   professional-license-ACC take-want-HDG ‖
          'OK, next, (I) wanna take a professional (boxing) license, or something like that,'
      b. (I) started to think like this.   (S01M0182: 251.43-257.40)

S01M0182_license

- Exceptions
    - Pitch range is smaller than that of predicate because:
    - Elements are **given**
    - Unclear cases

6.3 Discussion

Experimental study (Nakagawa, 2011)

- Predicate-focus context

(29) Yesterday the speaker and his/her friend found an abondoned puppy on the street. The speaker brought it to his/her home. Today, the speaker tells the friend what happened to the puppy.

    sooieba   [**koinu**]$_T$ [**yuzut-ta**]$_F$-yo
    by.the.way puppy   give-PAST-FP
    'By the way, (I) gave the puppy (to somebody).'

- All-focus context

21

(30)   The speaker and his/her friend are working in an animal shelter. The friend was
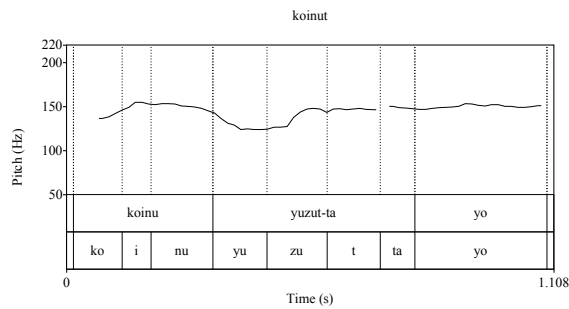   absent yesterday and wants to know what happened yesterday.

> kinoo-wa   [**koinu yuzut-ta**]$_F$-yo
> yesterday-TOP puppy give-PAST-FP
> 'Yesterday (we) gave puppies.'

- Predicate-focus context

(31)   Yesterday the speaker and his/her friend found an abondoned puppy on the street.
        The speaker brought it to his/her home. Today, the speaker tells the friend what
        happened to the puppy.

> sooieba   [**koinu**]$_T$ [**yuzut-ta**]$_F$-yo
> by.the.way puppy   give-PAST-FP
> 'By the way, (I) gave the puppy (to somebody).'

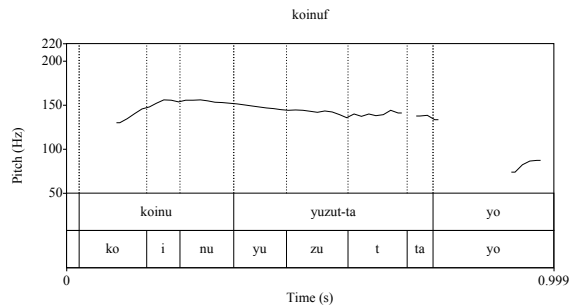- Pitch reset at the first mora of the predicate

- All-focus context

(32)   The speaker and his/her friend are working in an animal shelter. The friend was
        absent yesterday and wants to know what happened yesterday.

> kinoo-wa   [**koinu yuzut-ta**]$_F$-yo
> yesterday-TOP puppy give-PAST-FP
> 'Yesterday (we) gave puppies.'

- **No** Pitch reset at the first mora of the predicate

22

Summary

- Findings
    - A unit of IS corresponds to an IU
    - An element of low activation cost cannot form an IU alone

Discussion

- The iconic principle of intonation unit and information structure

    (33)   In spoken language, an IU tends to correspond to a unit of IS.

- The principle of intonation unit and activation cost

    (34)   all substantive IUs have similar activation costs; there are few IUs with only a strongly active element or those with too much new elements.

Discussion

- Principle of the separation of reference and role (Lambrecht, 1994)

    (35)   a. | Topic | ‖
            b. | Clause1 | ‖
            c. | Clause2 | ‖
            d. | Clause3 | ‖
            e.   ...

7   Discussion

Summary   Table 9 & 10

- Proposal
    - **Multi-dimensional analysis** of IS
    - Methodology of cross-linguistic annotation

- From-old-to-new principle

    (36)   In languages in which word order is relatively free, the unmarked word order of constituents is old, predictable information first and new, unpredictable information last.   (Kuno (1978, p. 54), Kuno (2004, p. 326))

- Information-structure continuity principle

    (37)   A unit of information structure must be continuous in a clause; i.e., elements which belong to the same unit are adjacent with each other.

Table 9: Summary of topic

Activation status	Particles	Word order	Intonation
Strongly active	(Zero pronoun)		
	toiuno-wa, wa, Ø	Post-predicate	Clausal IU
Active		Clause-initial	Phrasal IU
Semi-active inferrable	*wa*, Ø		
Semi-active decining	COP-*kedo/ga*, Ø		
Inactive unused			
Inactive brand-new	–	–	–

Table 10: Summary of (broad) focus

	Particles	Word order	Intonation
A	*ga*	Pre-predicate	Clausal IU
Agent S	*ga*		
Patient S	*ga*, Ø		
P	Ø		

- Persistent-element-first principle

  (38) In languages in which word order is relatively free, the unmarked word order of constituents is persistent element first and non-persistent element last.

- Iconic principle of intonation unit and information structure

  (39) In spoken language, an IU tends to correspond to a unit of information structure.

- Principle of intonation unit and activation cost

  (40) all substantive IUs have similar activation costs; there are few IUs with only a strongly active element or those with too much new elements.

Competing motivations

- Multi-dimensional analysis of IS is compatible with the idea of "**competing motivations**" (Du Bois, 1985)

- or "**seepage**" (Comrie, 1979)

Soft vs. hard constraints

- Bresnan, Dingare, and Manning (2001, p. 29)

  - "**soft constraints mirror hard constrains**";
  - "[t]he same categorical phenomena which are attributed to hard grammatical constraints in some languages continue to show up as statistical preferences in other languages, motivating a grammatical model that can account for soft constraints"
  - See also Givón (1979); Bybee and Hopper (2001).

- Elements integrated into the predicate

  - Pronominal affixation
  - Noun incorporation

- Elements separated from the predicate

  - in some languages, indefinite non-generic NPs cannot in general be the subject; they can only be the subject of existential constructions (Givón, 1976, p. 173ff.)
  - the connection between the subject (A and S) and topic is strong and non-topical subjects are not allowed

8 Conclusion

Remaining issues

- Predication or judgement types

- Genres

9   References

Birner, B. J., & Ward, G. (1998). *Information status and noncanonical word order in English.* Amsterdam/Philadelphia: John Benjamins.

Birner, B. J., & Ward, G. (2009). Information structure and syntactic structure. *Language and Linguistics Compass*, *3*, 1167–1187.

Bolinger, D. (1965). *Forms of English.* MA: Harvard University Press.

Bresnan, J., Dingare, S., & Manning, C. (2001). Soft constraints mirror hard constraints: voice and person in English and Lummi. In M. Butt & T. H. King (Eds.), *Proceedings of the LFG 01 conferene* (pp. 13–32). CA.

Büring, D. (2007). Intonation, semantics and information structure. In G. Ramchand & C. Reiss (Eds.), *The oxford handbook of linguistic interfaces* (pp. 445–474). Oxford: Oxford University Press.

Bybee, J., & Hopper, P. (Eds.). (2001). *Frequency and the emergence of linguistic structure.* Amsterdam: John Benjamins.

Calhoun, S., Nissim, M., Steedman, M., & Brenier, J. (2005). A framework for annotating information structure in discourse. In A. Meyers (Ed.), *Proceedings of the workshop on frontiers in corpus annotations ii: Pie in the sky* (pp. 45–52). Ann Arbor.

Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subject, topics, and point of view. In C. N. Li (Ed.), *Subject and topic* (pp. 25–53). New York: Academic Press.

Chafe, W. L. (1994). *Discourse, consciousness, and time.* Chicago/London: Chicago University Press.

Chiarcos, C., Fiedler, I., Grubic, M., Hartmann, K., Ritz, J., Schwarz, A., et al. (2011). Information structure in African languages: corpora and tools. *Language Resources and Evaluation*, *45*, 361–374.

Chomsky, N. (1965). *Aspects of the theory of syntax.* MA: MIT Press.

Comrie, B. (1979). Definite and animate direct objects: a natural class. *Linguistica Silesiana*, *3*, 13-21.

Comrie, B. (1983). Markedness, grammar, people, and the world. In F. R. Eckman, E. A. Moravicsik, & R. Wirth Jessica (Eds.), *Markedness* (p. 85-106). New York/London: Premium Press.

Cowan, N. (2000). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185.

Cowan, N. (2005). *Working memory capacity.* New York: Psychology Press.

Du Bois, J. W. (1985). Competing motivations. In J. Haiman (Ed.), *Iconicity in syntax* (p. 343-366). Amsterdam: John Benjamins.

Endo, Y. (2014). *Nihongo kaatogurafii josetsu.* Tokyo: Hituzi. ((Introduction to the Cartography of Japanese Syntactic Structures))

Erteschik-Shir, N. (1997). *The dynamics of focus structure.* Cambridge: Cambridge University Press.

Erteschik-Shir, N. (2007). *Information structure: The syntax-discourse interface.* Oxford: Oxford University Press.

Firbas, J. (1975). On the thematic and non-thematic section of the sentence. In H. Ringbom (Ed.), *Style and text: Studies presented to nils erik enkvist* (pp. 317–334). Stockholm: SprÂkfrlaget Skriptor AB.

Fry, J. (2001). *Ellipsis and Wa-marking in Japanese conversation.* Unpublished doctoral dissertation, Stanford University, CA.

Fujii, Y. (1995). Nihongo-no gojun-no gyakuten-nitsuite: kaiwa-no naka-no jôhô-no nagare-o chûsin-ni. In K.-I. Takami (Ed.), *Nichieego-no uhooten'i-koobun* (p. 167-198). Tokyo: Hituzi. ((On word order inverstion in Japanese: flow of information in conversation))

Givón, T. (1976). Topic, pronoun, and grammatical agreement. In C. N. Li (Ed.), *Subject and topic* (p. 149-187). New York: Academic Press.

Givón, T. (1979). *On understanding grammar.* New York: Academic Press.

Givón, T. (Ed.). (1983). *Topic continuity in discourse.* Amsterdam/Philadelphia: John Benjamins.

Götze, M., Weskott, T., Endriss, C., Fiedler, I., Hinterwimmer, S., Petrova, S., et al. (2007). Information structure. In S. Dipper, M. Götze, & S. Skopeteas (Eds.), *Information structure in cross-linguistic corpora: annotation guidelines for phonology, morphology, syntax, semantics and information structure* (pp. 147–187). Potsdam: Universitätsverlag Potsdam.

Gundel, J. (1974). *The role of topic and comment in linguistic theory.* Unpublished doctoral dissertation, University of Texas at Austin.

Hajičová, E., Panevová, J., & Sgall, P. (2000). *A manual for tectogrammatical tagging of the Prague Dependency Treebank* (Tech. Rep.). ÚFAL/CKL. ((TR-2000-09))

Halliday, M. A. K. (1967). Notes on transitivity and theme in English, part II. *Journal of Linguistics*, *3*, 199-244.

Ishihara, S. (2011). Japanese focus prosody revisited: Freeing focus from prosodic phrasing. *Lingua*, *121*, 1870–1889.

Iwasaki, S. (1993). The structure of the intonation unit in Japanese. *Japanese/Korean Linguistics*, *3*, 39–53.

Jackendoff, R. (1972). *Semantic interpretation in generative grammar.* MA: MIT Press.

Keenan, E. L. (1976). Towards a universal definition of "subject". In C. N. Li (Ed.), *Subject and topic* (p. 303-334). New York: Academic Press.

Kikuchi, Y. (1995). *Wa-kôbun-no gaikan.* In T. Masuoka, H. Noda, & Y. Numata (Eds.), *Nihongo-no shudai-to toritate* (pp. 37–69). Tokyo: Kurosio. ((Notes on narrative *wa*))

Kinsui, S. (1995). "katari-no *wa*"-ni kansuru oboegaki. In T. Masuoka, H. Noda, & Y. Numata (Eds.), *Nihongo-no shudai-to toritate* (p. 71-80). Tokyo: Kurosio. ((Notes on narrative *wa*))

Kori, S. (2011). Tôkyô hôgen-niokeru hiroi fôkasu-no onsei-teki tokuchô: renzoku suru 2-go-ni fôkasu-ga aru baai. *Onsei Gengo-no Kenkyu*, *5*, 13–20. ((Phonetic characteristics of broad focus in Tokyo dialect: with two foci in a row))

Krifka, M., & Misan, R. (Eds.). (2012). *The expression of information structure.* Berlin/New York: Mouton de Gruyter.

Kruijff-Korbayová, I., & Steedman, M. (2003). Discourse and information structure. *Journal of Logic, Language and Information*, *12*, 249-259.

Kuno, S. (1973). *The structure of the Japanese language.* MA: MIT Press.

Kuno, S. (1978). *Danwa-no bumpô.* Tokyo: Taishukan. ((Grammar of Discourse))

Kuno, S. (2004). Empathy and direct discourse perspectives. In R. Horn Laurence & G. Ward (Eds.), *The handbook of pragmatics* (pp. 315–343). Oxford: Blackwell.

Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus and the mental representations of discourse referents.* Cambridge: Cambridge University Press.

Maekawa, K. (2003). *Corpus of Spontaneous Japanese*: Its design and evaluation. In *Proceedings of the ISCA and IEEE workshop on Spontaneous speech processing and recognition* (pp. 7–12). Tokyo.

Maekawa, K., Kikuchi, H., & Tsukahara, W. (2004). Corpus of Spontaneous Japanese: design,

annotation and XML representation. In *Proceedings of the international symposium on large-scale knowledge resources (LKR2004)* (pp. 19–24). Tokyo.

Masuoka, T. (2000). *Nihongo bumpô-no shosô.* Tokyo: Kurosio. ((Aspects of Japanese Grammar))

Masuoka, T. (2012). Zokusei-jojutsu-to shudai-hyôshiki: Nihongo-kara-no apurôchi. In T. Kageyama (Ed.), *Zokusei-jojutsu-no sekai* (pp. 91–109). Tokyo: Kurosio. ((Property predication and topic marker: an approach from Japanese))

Mathesius, V. (1928). On linguistic characterology with illustrations from modern English. In J. Vachek (Ed.), *A Prague School reader in linguistics* (pp. 59–67). IN: Indiana University Press.

Mathesius, V. (1929). Functional linguistics. In J. Vachek (Ed.), *Praguiana: some basic and less well known aspects of the Prague Linguistics School* (p. 121-142). Amsterdam: John Benjamins.

Matsuda, K. (1996). *Variable zero-marking of (o) in Tokyo Japanese.* Unpublished doctoral dissertation, University of Pennsylvania, PA.

Matsumoto, K. (2000). Intonation units, clauses and preferred argument structure in conversational Japanese. *Language Sciences*, *22*, 63–86.

Matsushita, D. (1928). *Kaisen hyoôjun nihon bumpô.* Tokyo: Kigensha. ((New Basic Japanese Grammar))

Mikami, A. (1953/1972). *Gendei-go-hô josetsu.* Tokyo: Kurosio. ((Introduction to Modern Japanese Grammar))

Mikami, A. (1960). *Zô-wa hana-ga nagai.* Tokyo: Kurosio. ((The Elephant, the Nose is Long))

Mithun, M. (1995). Morphological and prosodic forces shaping word order. In P. Downing & M. Noonan (Eds.), *Word order in discourse* (pp. 387–423). Amsterdam/Philadelphia: John Benjamins.

Nakagawa, N. (2011). Hatuwa-no "mizikai tan'i"-o kôsei suru dôki-zuke: jôhô-kôzô-no kanten-kara. In *Slud* (p. 11-16). ((Information structural motivations for Short-Utterance Units))

Nakagawa, N., Asao, Y., & Nagaya, N. (2008). Information structure and intonation of right-dislocation sentences in Japanese. *Kyoto University Linguistic Research*, *27*, 1-22.

Nakagawa, N., Yokomori, D., & Asao, Y. (2010). The short intonation unit as a vehicle of important topics. *Papers in Linguistic Science*, *15*, 111-131.

Noda, H. (1996). *Wa to Ga* (Vol. 1). Tokyo: Kurosio. ((*wa* and *ga*))

Ono, T. (2007). An emotively motivated post-predicate constituent order in a 'strict predicate final' language: emotion and grammar meet in Japanese everyday talk. In S. Suzuki (Ed.), *Emotive communication in japanese.* Amsterdam: John Benjamins.

Ono, T., & Suzuki, R. (1992). Word order variability in Japanese conversation: motivations and grammaticalization. *Text*, *12*(3), 429-445.

Onoe, K. (1981). *Wa*-no kakari-joshi-sei-to hyôgen-teki kinô. *Kokugo-to Kokubun-gaku*, *56*(5), 102–118. ((*Kakari-joshi-hood* and expressive functions of *wa*))

Prince, E. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (p. 223-256). New York: Academic Press.

Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (Ed.), *Elements of grammar* (p. 281-337). Netherlands: Kluwer Academic Publishers.

Rooth, M. (1985). *Association with focus.* Unpublished doctoral dissertation, University of Massachusetts, Massachusetts.

Russell, B. (1905). On denoting. *Mind*, *14*, 479–493.

Saito, M. (1985). *Some asymmetries in Japanese and their theoretical consequence.* Unpublished

doctoral dissertation, Massachusetts Institute of Technology, Cambridge.

Selkirk, E. (1984). *Phonology and syntax: the relation between sound and structure*. MA: MIT Press.

Sgall, P. (1967). Functional sentence perspective. *Prague Studies in Mathematical Linguistics*, *2*, 203-225.

Steedman, M. (1991). Structure and intonation. *Language*, *67*, 262–296.

Strawson, P. F. (1950). On referring. *Mind*, *59*, 320–344.

Strawson, P. F. (1964). Identifying reference and truth values. *Theoria*, *30*, 86–99.

Tokieda, M. (1950/2005). *Nihon bumpô kôgo-hen*. Tokyo: Iwanami. ((Colloquial Japanese Grammar))

Tomlin, R. S. (1986). *Basic word order: Functional principles*. New Hampshire: Croom Helm.

Tsutsui, M. (1984). *Particle ellipses in Japanese*. Unpublished doctoral dissertation, University of Illinois, Urbana, Illinois.

Vallduví, E. (1990). *The information component*. Unpublished doctoral dissertation, University of Pennsylvania.

Vallduví, E., & Vilkuna, M. (1998). On rheme and kontrast. In P. W. Culicover & L. McNally (Eds.), *The limits of syntax* (pp. 79–108). San Diego: Academic Press.

Yamada, Y. (1936). *Nihon bumpô-gaku gairon*. Tokyo: Hobunkan. ((A Basic Theory of Japanese Grammar))