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, toino-wa, and kedo/ga preceded by copula) are mainly sensitive to activation status, whereas case particles (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-predicatively, 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 3
2 Background 4
3 Framework 5
4 Particles 7
5 Word order 12
6 Intonation 17
7 Discussion 23
8 Conclusion 25
9 References 26
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.’

2. B2: un tanosii-yo:
   yes fun-FP
   ‘Yeah, (it’s) fun.’

   good-FP a.bit  music-club-DAT enter-want-PAST-FP

4. C4: ima-kara-de-mo **gassyoosoo-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)
  - **Both traditions** (Vallduví, 1990; Steedman, 1991; Vallduví & Vilkuna, 1998)

- Roots of studies on IS
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 study
  - 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

<table>
<thead>
<tr>
<th>topic</th>
<th>focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. presupposed</td>
<td>asserted</td>
</tr>
<tr>
<td>b. active</td>
<td>inactive</td>
</tr>
<tr>
<td>c. definite</td>
<td>indefinite</td>
</tr>
<tr>
<td>d. specific</td>
<td>non-specific</td>
</tr>
<tr>
<td>e. animate</td>
<td>inanimate</td>
</tr>
<tr>
<td>f. agent</td>
<td>patient</td>
</tr>
<tr>
<td>g. inferable</td>
<td>non-inferable</td>
</tr>
</tbody>
</table>

(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 uncontroversional in a given sentence both by the speaker and the hearer.

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

- Uncontroversional: 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 controversional. 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

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

Table 2: Activation status in the corpus

<table>
<thead>
<tr>
<th>Activation status</th>
<th>Given-new taxonomy</th>
<th>Corpus annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>Evoked</td>
<td>Given</td>
</tr>
<tr>
<td>Semi-active</td>
<td>Declining</td>
<td></td>
</tr>
<tr>
<td>Semi-active</td>
<td>Inferable</td>
<td></td>
</tr>
<tr>
<td>Inactive</td>
<td>Unused</td>
<td>New</td>
</tr>
<tr>
<td>Inactive</td>
<td>Brand-new</td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Activation status</th>
<th>Given-New taxonomy</th>
<th>Topic</th>
<th>Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly active</td>
<td>Evoked</td>
<td>(Zero pronoun)</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Overt pronoun)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>toiuno-wa, wa, Ø</td>
<td></td>
</tr>
<tr>
<td>Active</td>
<td>Evoked</td>
<td>toiuno-wa, wa, Ø</td>
<td>case markers, Ø</td>
</tr>
<tr>
<td>Semi-active</td>
<td>Inferable</td>
<td>wa, Ø</td>
<td></td>
</tr>
<tr>
<td>Semi-active</td>
<td>Declining</td>
<td>COP-kedo/ga, Ø</td>
<td></td>
</tr>
<tr>
<td>Inactive</td>
<td>Unused</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inactive</td>
<td>Brand-new</td>
<td></td>
<td>–</td>
</tr>
</tbody>
</table>

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 un-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 sekai-taitoru-sen-o-desu-ne ee terebi-de mi-masi-ta
    FL world-title-fight-ơ-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
‘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
(7)  
a. There is a dish called chelow kebab.

b. de sore-wa ceto gohan-ni ceto 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 nyuusya 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 ii-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’
  – tinomogo ‘infant’

copula + kedo/ga

• Semi-active declining elements

(11) a. kore-kara ano mokuhyoo-tteiuno-ga arimasi-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 meesee-no bubun-to sigoto-tteiu bubun-ga arimasi-te FL fame-GEN part-and job-called part-ga exist-PLT-and
     ‘I have two goals: one is for fame and the other is for job.’
  d. Concerning fame,
  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 thii-taimu-tte 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,’
Figure 3: Anaphoric distance vs. expression type (all)

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

d. ee kanari ee soui 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-Ø 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-FF
   ‘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
Table 4: Overt vs. zero case markers

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>S</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Contrastive Focus</td>
<td>ga</td>
<td>ga</td>
<td>ga/Ø</td>
</tr>
<tr>
<td>Contrastive Focus</td>
<td>ga</td>
<td>ga</td>
<td>o</td>
</tr>
</tbody>
</table>

Table 5: Word order vs. activation status

<table>
<thead>
<tr>
<th>Activation status</th>
<th>Given-New taxonomy</th>
<th>Topic</th>
<th>Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly active (Zero pronoun)</td>
<td>Evoked</td>
<td>(Zero pronoun)</td>
<td>Pre-predicate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Overt pronoun)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-predicate</td>
<td></td>
</tr>
<tr>
<td>Active</td>
<td>Evoked</td>
<td>Clause-initial</td>
<td></td>
</tr>
<tr>
<td>Semi-active</td>
<td>Inferable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semi-active</td>
<td>Declining</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inactive</td>
<td>Unused</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inactive</td>
<td>Brand-new</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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
      ‘(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-ritten-nos o 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. totyu totyu-de-no 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
(16) a. ‘Also for Kilauea, (we) bought a map and’
   b. de zibun-tati-de ma rentakaa kuruma-o tobasi-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 tome-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 do-PROG-NMLZ this person
   ‘What is (he) doing, this person?’  (D02F0028: 193.30–194.45)

(18) L: sangurasu-toka kake-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)
Table 6: RD of post-predicate elements

<table>
<thead>
<tr>
<th>Single-contour</th>
<th>Double-contour</th>
</tr>
</thead>
<tbody>
<tr>
<td>RD</td>
<td>6.9</td>
</tr>
</tbody>
</table>

- 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)

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>RD</td>
<td>20.9</td>
<td>23.0</td>
</tr>
</tbody>
</table>
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 moti 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 si-masi-te mysterious impression-ga do-PLT-and ‘the impression was mysterious.’ (S00F0014: 1042.88–1047.03)

• The tendency holds regardless of word order
5.3 Discussion

Discussion

- Findings

\[(21)\]

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

- 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.

6 Intonation

Phrasal vs. clausal IUs

- Dependent variables

\[(23)\]

\(a. \quad \text{Phrasal IU}: \begin{array}{c} \text{NP} \\ \text{Predicate} \end{array}\)
\(b. \quad \text{Clausal IU}: \begin{array}{c} \text{NP} \\ \text{Predicate} \end{array}\)

Summary Table 8

Results Figure 16 & 17
Table 8: Intonation vs. activation status

<table>
<thead>
<tr>
<th>Activation status</th>
<th>Given-New taxonomy</th>
<th>Topic</th>
<th>Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly active</td>
<td>Evoked</td>
<td>(Zero pronoun)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Clausal IU</td>
<td>–</td>
</tr>
<tr>
<td>Active</td>
<td>Evoked</td>
<td>Phrasal IU</td>
<td>Clausal IU</td>
</tr>
<tr>
<td>Semi-active</td>
<td>Inferable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semi-active</td>
<td>Declining</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inactive</td>
<td>Unused</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inactive</td>
<td>Brand-new</td>
<td></td>
<td>–</td>
</tr>
</tbody>
</table>

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 F₀ peak
- Especially pronouns are cliticized.
- Element and predicate form a single processing unit.

\[(26)\]

```
(26) sore-wa || nan-daroo-to || omot-te ||
    that-wa || what-COP.INFR-QUOT think-and ||
    ‘(I) was wondering what it was...’
```

\[(S00F0014: 654.06-655.18)\]

6.2 Focus

Foci in clausal IUs

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

\[(27)\]

```
(27) a. our way of collecting debt might be problematic,
    b. oo mina-san || zisyuku suru-yoomi-to in || o-hanasi-ga
       FL everyone-HON || control do-IMP-QUOT say || PLT-speech-NOM
       de-masi-te ||
       come.out-PLT-and ||
    ‘somebody proposed that employees should improve the method.’
    \[(S00M0221: 503.23-511.02)\]
```
• No pitch reset & no pause & no lengthening & no rising

(28) a. ano puro-raisensu-o tori-tai-toka ||
\[FL \ \text{professional-license-ACC} \ \text{take-want-HDG} ||
\]
‘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)

• 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 abandoned puppy on the street. The speaker brought it to his/her home. Today, the speaker tells the friend what happened to the puppy.

sooieba \[\text{koinu}\text{[T]} \ \text{[yuzut-ta]}\text{[FP]}-yo\]
by.the.way puppy give-PAST-FP
‘By the way, (I) gave the puppy (to somebody).’

• All-focus context
(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.

\[
\text{kinoo-wa } \{\text{koinu yuzut-ta}\}_F - \text{yo}
\]

yesterday-TOP puppy give-PAST-FP

‘Yesterday (we) gave puppies.’

- Predicate-focus context

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

\[
\text{sooieba } \{\text{koinu}\}_T \{\text{yuzut-ta}\}_F - \text{yo}
\]

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

(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.

\[
\text{kinoo-wa } \{\text{koinu yuzut-ta}\}_F - \text{yo}
\]

yesterday-TOP puppy give-PAST-FP

‘Yesterday (we) gave puppies.’

- No Pitch reset at the first mora of the predicate
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

<table>
<thead>
<tr>
<th>Activation status</th>
<th>Particles</th>
<th>Word order</th>
<th>Intonation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly active</td>
<td><em>(Zero pronoun)</em></td>
<td>Post-predicate</td>
<td>Clausal IU</td>
</tr>
<tr>
<td>toiuno-wa, wa, Ø</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active</td>
<td></td>
<td>Clause-initial</td>
<td>Phrasal IU</td>
</tr>
<tr>
<td>Semi-active</td>
<td>wa, Ø</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semi-active</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>decining</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inactive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unused</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inactive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>brand-new</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 10: Summary of (broad) focus

<table>
<thead>
<tr>
<th>Particles</th>
<th>Word order</th>
<th>Intonation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>ga</td>
<td>Pre-predicate</td>
</tr>
<tr>
<td>Agent S</td>
<td>ga</td>
<td></td>
</tr>
<tr>
<td>Patient S</td>
<td>ga, Ø</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>Ø</td>
<td></td>
</tr>
</tbody>
</table>
• 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


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


