A Corpus-Based Analysis of Canonical Word Order of Japanese Double Object Constructions

Ryohei Sasano and **Manabu Okumura** (Tokyo Institute of Technology)

1. Introduction

Japanese double object construction:

ケンが	アヤに	カメラを	見せた。
Ken-NOM	Aya-DAT	camera-ACC	showed

(Ken showed a camera to Aya.)

verb

There are several claims as to the dative **(DAT)** and accusative **(ACC)** order

Background: most theoretical or empirical studies required manual analyses or measurements of human characteristics for each example

Assumption: Relationship between the canonical word order and the proportion of each word order

ACC-DAT (99.6%): 愛情を (affection-ACC) 言葉に (word-DAT) 感じる。(feel)
(ϕ_I feel the affection in ϕ_{your} words.)

DAT-ACC (97.5%): デートに (date-DAT) 女性を (woman-ACC) 誘う。(ask)
(ϕ_I ask a woman out on a date.)

This study presents a corpus-based analysis of the canonical word order

2. Claims on the word order

Our analysis suggests

A) The DAT-ACC order is canonical (Hoji 1985) — No
B) There are two canonical word orders, the DAT-ACC and the ACC-DAT order, depending on the verb types (Matsuoka 2003) — No
C) An argument whose grammatical case is infrequently omitted with a given verb tends to be placed near the verb — Yes
D) Canonical word order varies depending on the semantic role and animacy of the dative argument (Matsuoka 2003) — Yes
E) An argument that frequently co-occurs with the verb tends to be placed near the verb — Yes

3. Example collection

- We extracted examples of verbs and their arguments from a corpus consisting of more than 10 billion Web sentences
- We used only unambiguous parts of dependency parses, and collected the verb that had more than 500 different examples

→ : other candidates
→ : dependency

e.g. カギを (key-ACC) 彼に (him-DAT) 言われた (told) 場所に (place-DAT) 置いた (put)。
(ϕ_I put the key on the place where he told *me*.)

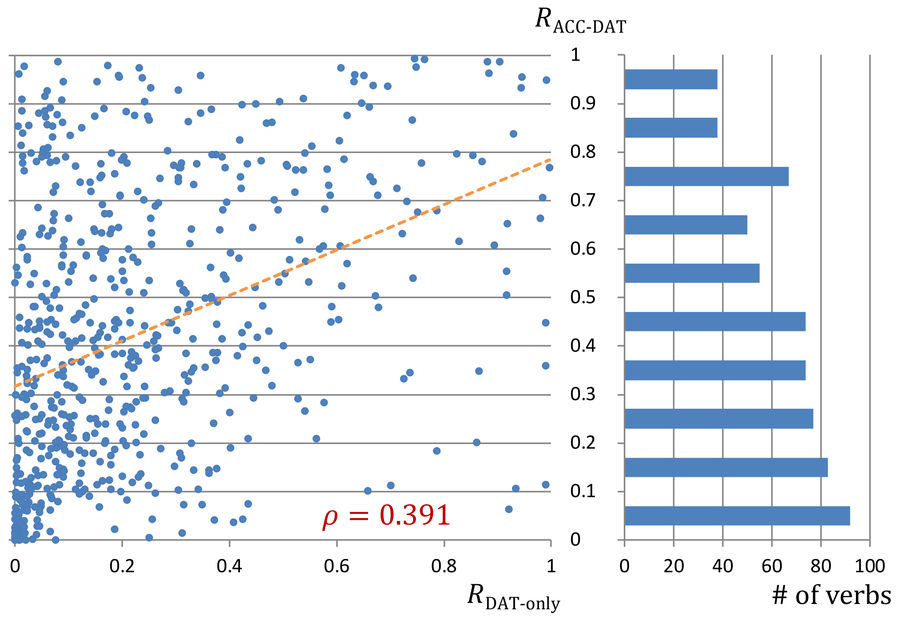
- ~~彼に 言われた~~
- ~~カギを~~
  場所に 置いた

4. Analysis

1. Word order for each verb (Claim A and Claim C) — No / Yes

- The relation between the proportion of DAT-only example $R_{\mathrm{DAT-only}}$ and that of the ACC-DAT order $R_{\mathrm{ACC-DAT}}$ for each collected 648 verbs

$$R_{\mathrm{ACC-DAT}} = \frac{N_{\mathrm{ACC-DAT}}}{N_{\mathrm{DAT-ACC}} + N_{\mathrm{ACC-DAT}}}, \quad R_{\mathrm{DAT-only}} = \frac{N_{\mathrm{DAT-only}}}{N_{\mathrm{DAT-}only} + N_{\mathrm{ACC-DAT}}}$$



2. Word order and verb type (Claim B) — No

- There is a claim that the DAT-ACC order is canonical for show-type verbs, whereas the ACC-DAT order is canonical for pass-type verbs

Show-type		Pass-type			
Verb	$R_{\mathrm{ACC-DAT}}$	verb	$R_{\mathrm{ACC-DAT}}$	Verb	$R_{\mathrm{ACC-DAT}}$
知らせる(notify)	0.522	戻す(put back)	0.771	落とす(drop)	0.351
預ける(deposit)	0.399	泊める(lodge)	0.748	漏らす(leak)	0.332
事付ける(request)	0.386	包む(wrap)	0.603	浮かべる(float)	0.255
悟す(adomish)	0.325	伝える(inform)	0.522	向ける(direct)	0.251
見せる(show)	0.301	載せる(place on)	0.496	残す(leave)	0.238
被せる(cover)	0.256	届ける(deliver)	0.491	埋める(bury)	0.223
教える(teach)	0.235	並べる(range)	0.481	混ぜる(blend)	0.200
授ける(give)	0.186	返す(give back)	0.448	当てる(hit)	0.185
浴びせる(shower)	0.177	ぶつける(knock)	0.436	掛ける(hang)	0.108
貸す(lend)	0.118	付ける(attach)	0.368	重ねる(pile)	0.084
着せる(dress)	0.113	渡す(pass)	0.362	建てる(build)	0.069
Macro average	0.274			Macro average	0.367

3. Word order and semantic role (Claim D) — Yes

- There is a claim that the ACC-DAT order is more preferred when the semantic role of the DAT is inanimate **Goal** than when the role is animate **Possessor**

本を (book-ACC) 学校に (school-DAT) 返した。(returned)
(ϕ returned the book to school.)

⇔

先生に (teacher-DAT) 本を (book-ACC) 返した。(returned)
(ϕ returned the book to the teacher.)

- We collected the examples that satisfied the following conditions:
  - **A)** ACC=ARTIFACT & DAT=**PLACE-INSTITUTION**
  - **B)** ACC=ARTIFACT & DAT=**PERSON**
- We extracted verbs that had at least 100 examples of both types
- Out of 126 verbs, 64 verbs show the trend that **Type-A** prefers the ACC-DAT order more than **Type-B** does, and only 30 verbs have the opposite trend

4. Word order for each tuple of a verb and arguments (Claim E) — Yes

- We examined the relation between $R_{\mathrm{ACC-DAT}}$ and the degree of co-occurrence of a verb and its argument
- We investigated 2302 tuples of a verb and its arguments
- We used the NPMI for measuring the degree of co-occurrence

$$\mathrm{NPMI}(n_c, v) = \frac{\mathrm{PMI}(n_c, v)}{-\log(p(n_c, v))}$$

+/**x** denotes the verb and ACC/DAT argument are used as an idiom