What discourse analysis tells us about discourse-pragmatics in RRG

Mitsuaki Shimojo
shimojo@buffalo.edu
Department of Linguistics
University at Buffalo, The State University of New York

1. Introduction

2. RRG representations of discourse-pragmatics

3. Analysis I: topic and zero anaphora in L1, L2 Japanese written narratives & RRG application

4. Analysis II: topic in L1 Japanese and Korean written narratives & RRG application

5. Summary

Background

- Saliency (or topicality, continuity) of information in discourse influences speakers' pragmatic intentions with respect to the choice of sentence forms.

- Salient information tends to be represented by reduced forms, or even omitted, and marked placement such as dislocation (Prince 1992, Gundel, et al 1993, Chafe 1994, Fry 2003, Shimojo 2005).

Questions

- Discourse generalization: How do speakers choose one form over the other, with respect to topicalization and omission (zero anaphora)?

- RRG application: How are the discourse findings captured in RRG? How does pragmatics influence the RRG representations with respect to the use of the forms?

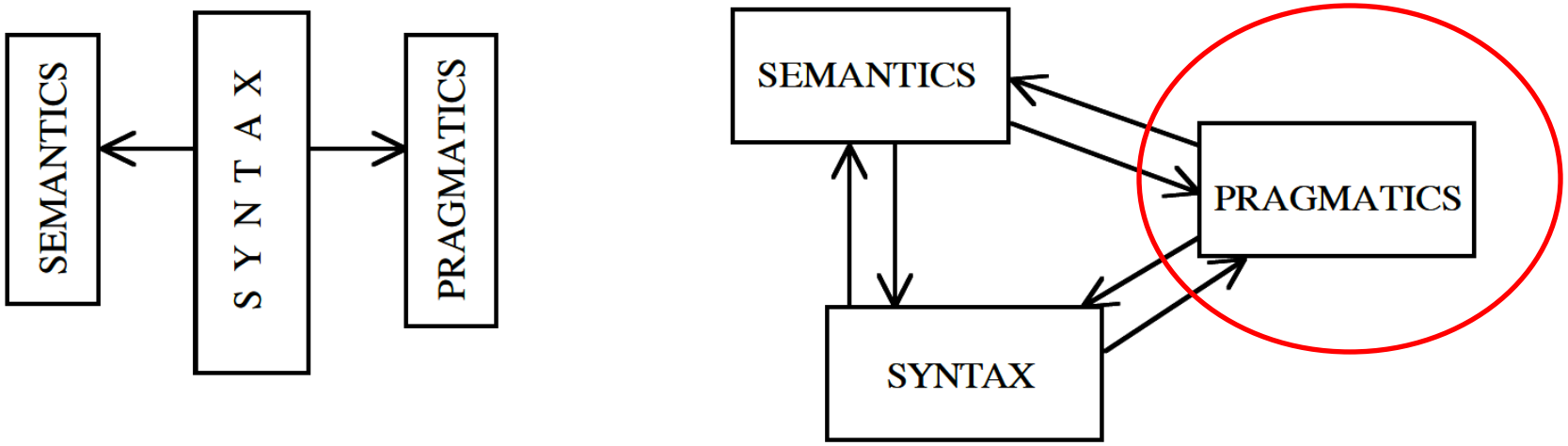


Figure 1 : Syntactocentric vs. parallel architecture theories

Van Valin (2014)

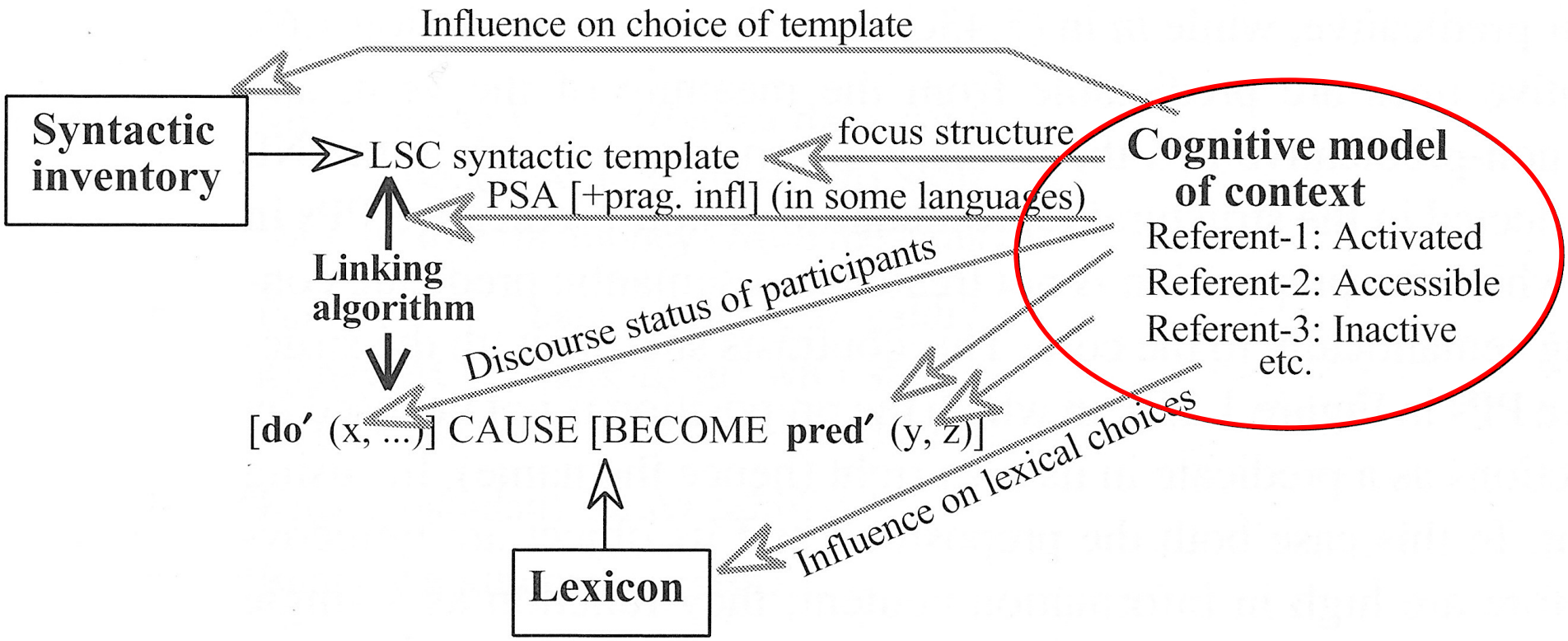


Figure 5.40 *The pervasive role of discourse-pragmatics in grammar*

Discourse-pragmatics "can play a role in virtually every aspect of grammar" (Van Valin 2005: 182)

Discourse Representation Structure
(referents, propositions, presuppositions, assertions)

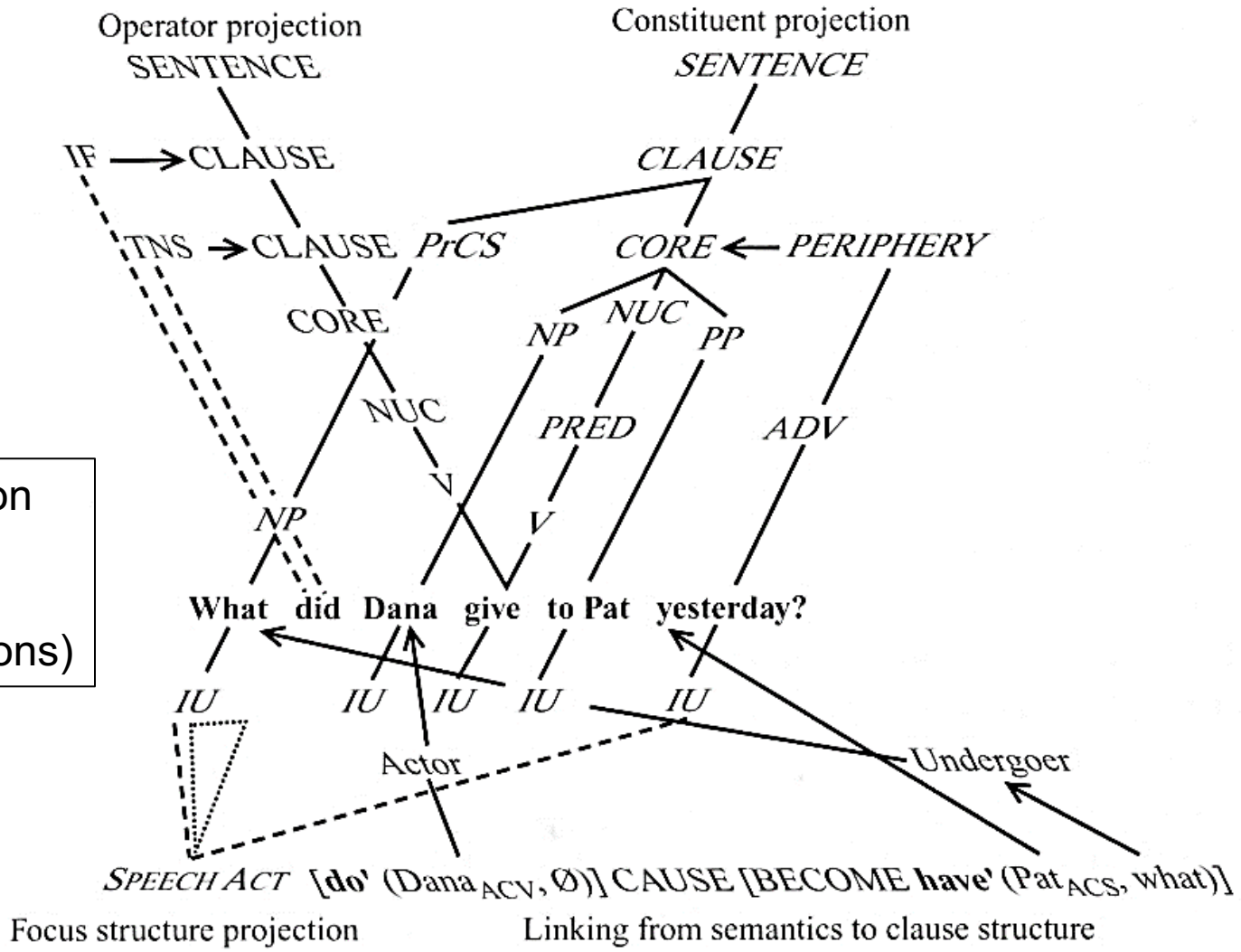


Figure 5.30 *Interaction of linking with full clause structure*
(Van Valin 2005)

Zero anaphora in Mandarin (Van Valin 2005: 174)

(5.33) a. Lǎo Qián$_i$ yǒu zhème ge píqì,
Old Qian have such CL disposition
'Old Qian$_i$ has (just) such a disposition:

b. *pro*$_i$ wèn péngyǒu$_j$ yào shénme dōngxi$_k$,
ask friend want what/something thing
if (he$_i$) asks for something from (his) friend(s),

c. *pro*$_j$ lìkè jiù děi gěi *pro*$_i$ *pro*$_k$.
at-once then must give
(he/she/they$_j$) must give (it$_k$) (to him$_i$) at once.'

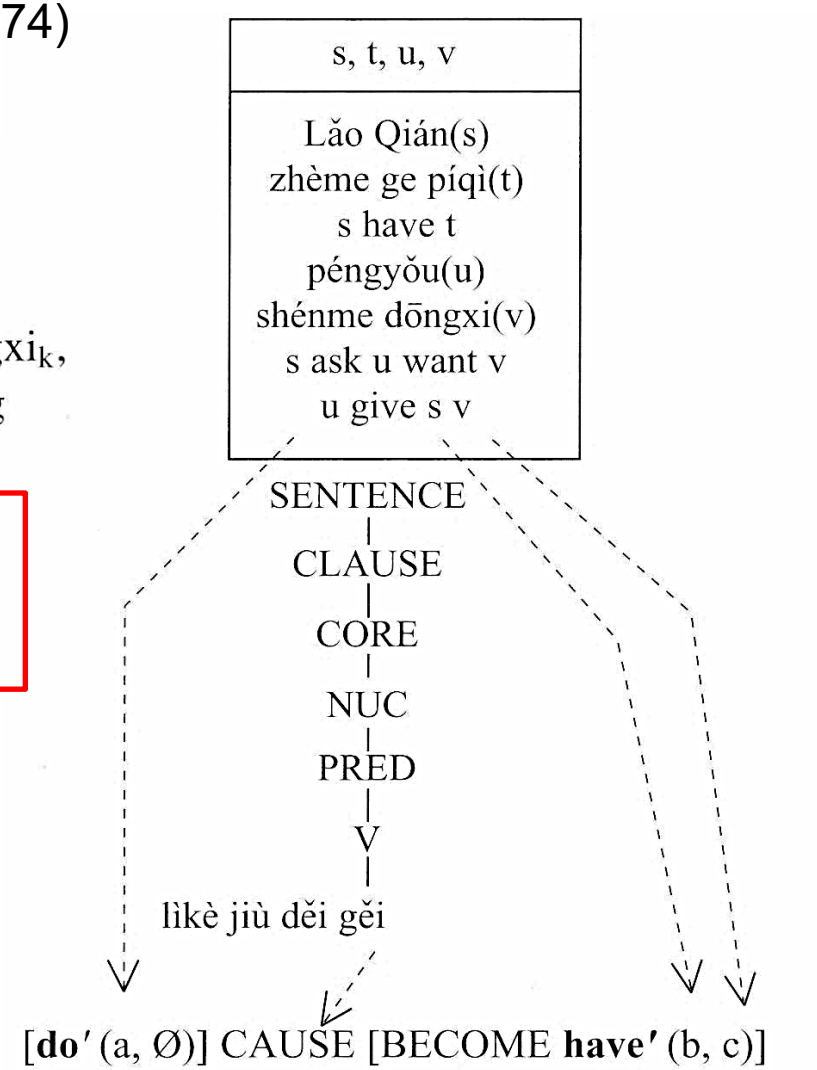


Figure 5.37 *Linking directly from discourse re*
structure to satisfy the completeness constrain

Topicalization and zero anaphora in Japanese

1. otokonohito-ga okusan-to hanashiteimasu. (P03)
   man-NOM wife-with talking
   'A man is talking with his wife.'
2. otoko-wa dekakeru junbi-o hajimemashita.
   man-TOP go.out preparation-ACC began
   'The man was getting ready to go out.'
3. okusan-o oite
   wife-ACC leave.and
   '(The man) leaves his wife (at home) and'
4. disuko-ni mukaimasu.
   disco-to head.for
   '(the man) heads for a disco.'

Topicalization: left dislocated in LDP and marked with the topic marker *WA*

Zero anaphora

The two forms are functionally similar.

- They represent previously-given, non-focal, and persistent information (e.g. Fry 2003, Hinds & Hinds 1979, Shimojo 2005, Suzuki 1995, Watanabe 1989).

- They are often interchangeable in a given context. Per Kuno (1973: 222), ellipted subjects are derived from *WA*-marked topics.

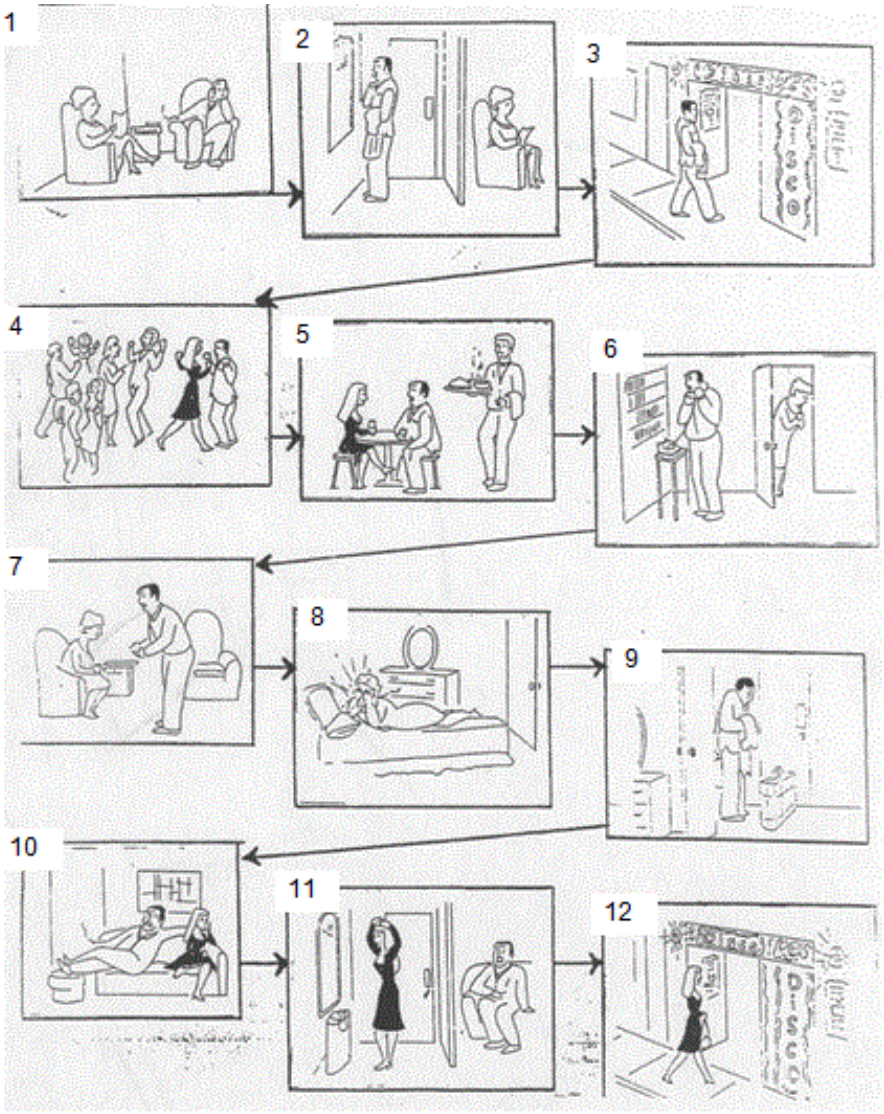
However, they are not functionally identical.

- What is omitted is not always a topic of the sentence (e.g. more than one argument of a sentence may be omitted).

- An overt topic is often used even if the intended referents are contextually identifiable.

Discourse data

(1) Picture-based narratives



Based on Brown & Yule (1983)

(2) The animation-based narratives

10

"Pingu Runs Away"

Subject groups

Proficiency: low ↕ high

Subject group	#	Japanese class enrolled at data collection	Narrative	Total # of clause
Native	11	--	picture	346
L1	62	--	animation	1557
Learners	3	2nd year-2nd semester	picture	80
L2-1	5	3rd year-1st semester	picture	125
Learners	2	3rd year-2nd semester	picture	44
L2-2	8		animation	245
Learners	11	4th year	animation	365
L2-3	13		animation	399

Learners are all L1 English speakers.

- The data from the picture and animation-based narratives were combined for analysis.
- Total number of clauses: L1 - 1903, L2 - 1258

Framework for analysis: Centering Theory
(Grosz & Sidner 1986, Grosz et al. 1995; Walker et al. 1998)

- Discourse segments exhibit both local coherence (among the utterances within a discourse segment) and global coherence (with other segments in the discourse).

- Centering Theory is concerned with local coherence and saliency, i.e. the discourse participants' focus of attention and how their attentional state is updated clause by clause.

- Defines different transition types of how information continues from one clause to the next.

- Each utterance has two structures of local focus: forward-looking centers [CFs] (currently talked-about entities) and a backward-looking center [CB] (the centrally talked-about entity or center of attention).

- All nominal referents represented by an utterance are CFs but only one of them may be the CB.

- CFs are ranked in terms of the likelihood to continue in the following utterance.

CF ranking for Japanese

(Overt) TOPIC > EMPATHY > SUBJECT > OBJECT2 (recipient of ditransitive) > OBJECT > OTHERS

*Based on Walker, Iida & Cote (1994) but slightly modified.
Omitted arguments are ranked according to their canonical roles.

The highest ranked CF is the preferred center [CP].

1. A man is talking with his wife. CP CF	[man, wife]
2. He was getting ready to go out. CP	[man]
3. He leaves his wife at home and CP CF CF	[man, wife, home]
4. ∅ heads for a disco. CP CF	[man, disco]

CB: centrally talked-about entity (center of attention)
CF of the current utterance (Ui) which is the highest ranked CF of the immediately preceding utterance (Ui-1).

- Five transition types

Cataphoric property	Anaphoric property: CB(Ui)=CB(Ui-1) Same CB	Anaphoric property: CB(Ui)≠CB(Ui-1) New CB
CB(Ui)=CP(Ui) Same center likely to continue in the next utterance.	CONTINUE	SMOOTH-SHIFT
CB(Ui)≠CP(Ui) Same center not likely to continue in the next utterance.	RETAIN	ROUGH-SHIFT
No CF in Ui is in Ui-1	NULL	

- Transition ordering rule (Grosz et al. 1995: 215)

Continue > Retain > Smooth-shift > Rough-shift

Smoother transition / greater coherence

CON is preferred to RET, which is preferred to S-SHIFT, which is preferred to R-SHIFT.

Analysis

WA-marked topics and omitted arguments were tagged as follows.

(1) Pragmatic properties

- Attention status (CB, non-CB)
- Transition types (CON, RET, S-SHIFT, R-SHIFT, NULL)

20. pinguu-wa CON kogoeteita node [Pingu]
    Pingu-TOP was.freezing because
    'Because Pingu was freezing'
21. okaasan-wa (Pingu-ni) RET moofu-o kaketeageta. [mother, Pingu]
    mother-TOP blanket-ACC put.over
    'the mother put the blanket over (Pingu).'
22. pinguu-wa CON otoosan-ni kuruma-ni nosetemoratte [Pingu, father, car]
    Pingu-TOP father-by car-in drive.and
    'Pingu was put in the car by the father.'
23. (Pingu-ga) CON ie-made kaetteitta [Pingu, home]
    home-to returned
    '(Pingu) went home.'

[L1: MJ05]

(2) Structural properties

- Placement of the clause: sentence-initial [SI], non-sentence-initial [~SI]
- Co-reference with the preceding (topicalized) subject: same subject [SS], different subject [DS]

20. pinguu-wa CON kogoeteita node [SI, DS]
Pingu-TOP was.freezing because
'Because Pingu was freezing'

21. okaasan-wa (Pingu-ni) RET moofu-o kaketeageta. [~SI, DS]
mother-TOP blanket-ACC put.over
'the mother put the blanket over (Pingu).'

22. pinguu-wa CON otoosan-ni kuruma-ni nosetemoratte [SI, DS]
Pingu-TOP father-by car-in drive.and
'Pingu was put in the car by the father.'

23. (Pingu-ga) CON ie-made kaetteitta [~SI, SS]
home-to returned
'(Pingu) went home.'

[L1: MJ05]

Overall coherence

Total number of clauses by transition type and SS/DS

	CON	RET	S-SHIFT	R-SHIFT	NULL	Total
L1	787	417	283	168	248	1903
L2	455	247	190	128	238	1258

L1 and L2 are consistent with the coherence ranking (per Grosz et al. 1995):
CON > RET > S-SHIFT > R-SHIFT

L2 used more NULL's; greater discontinuity of centers (less coherent).

CB forms (excluding NULL)

	L1	L2
Zero anaphora	895 (.54)	507 (.50)
WA	494 (.30)	349 (.34)
Other	266 (.16)	165 (.16)
Total	1655 (1.00)	1020 (1.00)

Zero anaphora is more frequent than *WA*.

Overuse of *WA* in L2

Zero anaphora and *WA* are most commonly used to represent center of attention.

CB and non-CB total

	L1	L2
Zero anaphora	1161 (.56)	643 (.49)
WA	920 (.44)	678 (.51)
Total	2081(1.00)	1321 (1.00)

Fisher's exact test: $p<.0001$

CB forms by transition type

L1	CON/S-SHIFT	RET/R-SHIFT
Zero anaphora	707 (.67)	189 (.57)
WA	348 (.33)	145 (.43)
Total	1055 (1.00)	334 (1.00)

Fisher's exact test: $p < .001$

L2	CON/S-SHIFT	RET/R-SHIFT
Zero anaphora	389 (.63)	118 (.49)
WA	224 (.37)	125 (.51)
Total	613 (1.00)	243 (1.00)

Fisher's exact test: $p < .0001$

Zero anaphora is more frequent in CON and S-SHIFT.

Cataphoric property	Anaphoric property: CB(Ui)=CB(Ui-1) Same CB	Anaphoric property: CB(Ui)≠CB(Ui-1) New CB
CB(Ui)=CP(Ui) Same center likely to continue in the next utterance.	CONTINUE zero anaphora > *wa*	SMOOTH-SHIFT zero anaphora > *wa*
CB(Ui)≠CP(Ui) Same center not likely to continue in the next utterance.	RETAIN zero anaphora, *wa*	ROUGH-SHIFT zero anaphora, *wa*

- Why CON and S-SHIFT?
  The CP (PSA) represents the center of attention (the center is highly salient because the two centers are represented by the same argument).
- "Saliency Ranking" [CON/S-SHIFT > RET/R-SHIFT] is more relevant to the use of zero anaphora than the coherence ranking [CON > RET > S-SHIFT > R-SHIFT].

CB: Zero anaphora and *WA* by transition type, SI/~SI, SS/DS

L2-1	CON-SS	SI	CON-SS	~SI	CON-DS	SI	CON-DS	~SI
Zero	27	0.54	25	0.86	2	0.40	0	
WA	23	0.46	4	0.14	3	0.60	0	
Total	50	1.00	29	1.00	5	1.00	0	
L2-2	**CON-SS**	**SI**	**CON-SS**	**~SI**	**CON-DS**	**SI**	**CON-DS**	**~SI**
Zero	13	0.35	50	0.93	2	0.25	1	0.17
WA	24	0.65	4	0.07	6	0.75	5	0.83
Total	37	1.00	54	1.00	8	1.00	6	1.00
L2-3	**CON-SS**	**SI**	**CON-SS**	**~SI**	**CON-DS**	**SI**	**CON-DS**	**~SI**
Zero	42	0.49	123	0.96	1	0.06	2	0.14
WA	44	0.51	5	0.04	17	0.94	12	0.86
Total	86	1.00	128	1.00	18	1.00	14	1.00
L1	**CON-SS**	**SI**	**CON-SS**	**~SI**	**CON-DS**	**SI**	**CON-DS**	**~SI**
Zero	85	0.38	406	0.93	8	0.12	16	0.38
WA	139	0.62	32	0.07	60	0.88	26	0.62
Total	224	1.00	438	1.00	68	1.00	42	1.00

L2-1, L1 p=.0365

For SS-SI, *WA* outnumbers in L1, but L2 is inconsistent, leaning more toward zero anaphora.

For SS-~SI, zero anaphora is dominant.

For DS (switched PSAs), *WA* is dominant.

CB: Zero anaphora and *WA* by transition type, SI/~SI, SS/DS

L2-1	S-SHIFT-SS	SI	S-SHIFT-SS	~SI	S-SHIFT-DS	SI	S-SHIFT-DS	~SI
Zero	6	0.30	5	0.71	0	0.00	0	0.00
WA	14	0.70	2	0.29	8	1.00	1	1.00
Total	20	1.00	7	1.00	8	1.00	1	1.00
L2-2	S-SHIFT-SS	SI	S-SHIFT-SS	~SI	S-SHIFT-DS	SI	S-SHIFT-DS	~SI
Zero	4	0.36	10	0.91	2	0.33	3	0.60
WA	7	0.64	1	0.09	4	0.67	2	0.40
Total	11	1.00	11	1.00	6	1.00	5	1.00
L2-3	S-SHIFT-SS	SI	S-SHIFT-SS	~SI	S-SHIFT-DS	SI	S-SHIFT-DS	~SI
Zero	13	0.59	53	0.93	2	0.10	3	0.30
WA	9	0.41	4	0.07	18	0.90	7	0.70
Total	22	1.00	57	1.00	20	1.00	10	1.00
L1	S-SHIFT-SS	SI	S-SHIFT-SS	~SI	S-SHIFT-DS	SI	S-SHIFT-DS	~SI
Zero	21	0.38	146	0.92	8	0.20	10	0.50
WA	34	0.62	13	0.08	33	0.80	10	0.50
Total	55	1.00	159	1.00	41	1.00	20	1.00

For SS-SI, inconsistency in L2 (also between CON and S-SHIFT)

For SS-~SI, zero anaphora is dominant.

Similar to CON overall.

CB: Zero anaphora and *WA* by transition type and SI/~SI

L2-1	RET	SI	RET	~SI	R-SHIFT	SI	R-SHIFT	~SI
Zero	9	0.45	2	0.67	6	0.40	2	0.67
WA	11	0.55	1	0.33	9	0.60	1	0.33
Total	20	1.00	3	1.00	15	1.00	3	1.00
L2-2	RET	SI	RET	~SI	R-SHIFT	SI	R-SHIFT	~SI
Zero	11	0.44	5	0.56	8	0.62	4	1.00
WA	14	0.56	4	0.44	5	0.38	0	0.00
Total	25	1.00	9	1.00	13	1.00	4	1.00
L2-3	RET	SI	RET	~SI	R-SHIFT	SI	R-SHIFT	~SI
Zero	35	0.46	18	0.67	9	0.29	9	0.53
WA	41	0.54	9	0.33	22	0.71	8	0.47
Total	76	1.00	27	1.00	31	1.00	17	1.00
L1	RET	SI	RET	~SI	R-SHIFT	SI	R-SHIFT	~SI
Zero	77	0.52	60	0.77	16	0.30	42	0.67
WA	71	0.48	18	0.23	37	0.70	21	0.33
Total	148	1.00	78	1.00	53	1.00	63	1.00

- RET and R-SHIFT are DS only.
- Mixed distribution; the choice depends on how the CB relates to the preceding clause.

WA in Retain and Rough Shift: partial coreference with CB/CF(Un-1)

(A19)

22. sannin-wa [SS] isshoni uchi-ni kaettekita. [Pingu family, home]
    3.CL-TOP together home-to returned
    'The three (Pingu family) came home together.'
23. pinguu-wa [R] uchi-ni tsuite kara, [Pingu family/Pingu, home]
    Pingu-TOP home-to arrive after
    'After Pingu arrived home'
24. (Pingu) [SS] oya-ni …suupu-o nomasetemorai, [Pingu, parents, soup]
    parents-by soup-ACC have.drink
    '(Pingu) had the parents give (him) …soup'
25. (Pingu) [C] karada-o atatameta. [Pingu, body]
    body-ACC warmed.up
    '(Pingu) warmed up (his) body.'

'family' (CB) to 'Pingu' (CP)
[whole to part]

(A19) cont.

26. pinguu-wa [C], jibun-no ryooshin-no yasashisa-o kanjitanodatta.
[Pingu, kindness, parents]

Pingu-TOP self-LK parent-LK kindness-ACC felt

'Pingu felt his own parents' kindness.'

27. sannin-wa [R] sono hi-no yoru, onaji beddo-de isshoni neta.
[Pingu/Pingu family, bed, night]

three.CL-TOP that day-LK night same bed-in together slept

'The three slept in the same bed together that night.'

'Pingu' (CB) to 'family' (CP)
[part to whole]

Zero anaphora in Retain and Rough Shift: identical coreference with CB/CF(Un-1)

[L1-MJ50]

1. pingu-wa yasai-ga kiraidatta [NULL]
   Pingu-TOP vegetable-NO hated
   'Pingu hated vegetables.'
2. aruhi yuushoku-ni hoorensoo-ga hukumareteita [NULL]
   one.day dinner-in spinach-NOM was.included
   'One day, there was spinach in the dinner.'
3. pingu-wa chichi-ni (spinach) [R] taberuyooni-iwareta ga
   Pingu-TOP father-by eat-was.told but
   'Pingu was told to eat (the spinach) by father but'
4. (Pingu) [SS] (spinach) tabenakatta.
   did.not.eat
   '(Pingu) did not eat (it).'

Zero anaphora:
identical coreference with
CF(Un-1)

RETAIN and R-SHIFT by co-reference types with CP(n-1)

L1	*WA*					zero anaphora				
	=	⊃	⊂	∩	≠	=	⊃	⊂	∩	≠
RET	2	32	51	0	4	103	25	1	0	8
R-SHIFT	5	15	29	1	8	21	20	1	0	16
Total	7	47	80	1	12	123	45	3	0	24

L2	*WA*					zero anaphora				
	=	⊃	⊂	∩	≠	=	⊃	⊂	∩	≠
RET	0	24	53	0	3	48	16	2	0	16
R-SHIFT	0	5	36	0	4	15	12	4	0	5
Total	0	29	89	0	7	63	28	6	0	22

⊃: CB(n) is a superset of CP(n-1), ⊂: CB(n) is a subset of CP(n-1),
∩: partial overlap

In both L1 and L2, *WA* outnumbers in partial coreference.

Identical coreference is typically represented by zero anaphora.
Also common when supersets are coreferential with preceding subsets. Assumption: the whole can be made accessible by a part more readily (without overt reference to the whole).

CB and non-CB

Zero anaphora	CB		non-CB		Total	
L1	895	(.77)	266	(.23)	1161	(1.00)
L2	507	(.79)	136	(.21)	643	(1.00)

Zero anaphora used primarily for center of attention.
Mostly nominative arguments.

Also used for non-CBs;
but commonly coreferential with U(n-1)
[66% in L1; 46% in L2]

WA	CB		non-CB		Total	
L1	494	(.54)	426	(.46)	920	(1.00)
L2	349	(.51)	329	(.49)	678	(1.00)

WA's is common across CB and ~CB (typically nominative arguments).
Non-CB's are coreferential with U(n-1) only in 14% (L1) and 13% (L2) of the total.

Summary of the findings

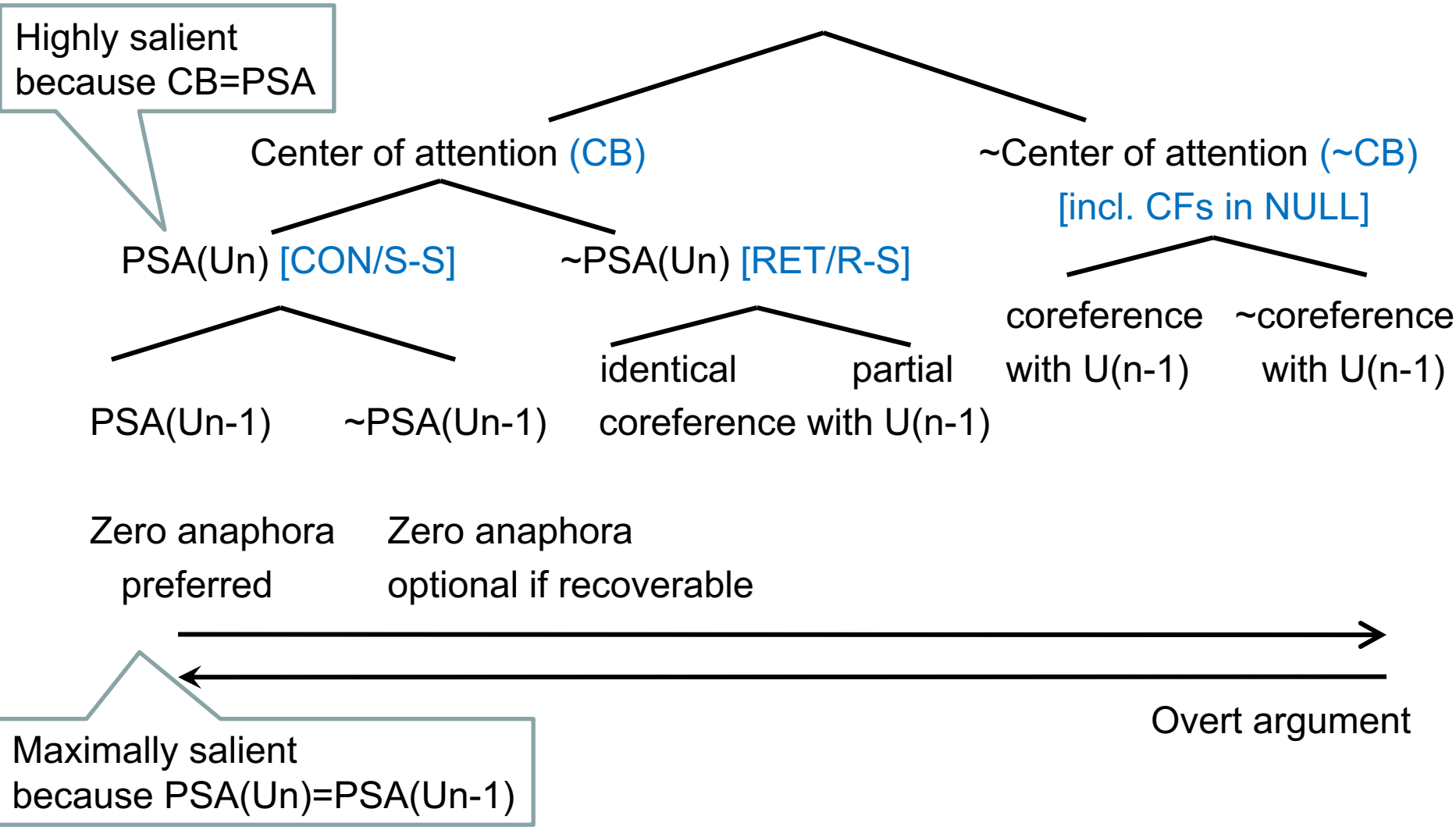
1. L1 and L2 narratives were consistent with the coherence ranking: CON > RET > S-SHIFT > R-SHIFT. Yet, the L2 speakers overused NULL, referentially discontinuous clauses.

2. Omission and topicalization of arguments are related to saliency of referents (not coherence in Centering). PSAs representing center of attention (i.e. CON and S-SHIFT) are highly salient and tend to be omitted.

3. Saliency is determined structurally as well. Highly salient referents are maximally salient when they are represented by a sequence of coreferential PSAs.

4. The use of an overt topic is independently related to a structural property; the PSA of a sentence-initial unit tends to be overtly expressed, at least in L1 (inconsistent in L2).

Saliency hierarchy and argument omission



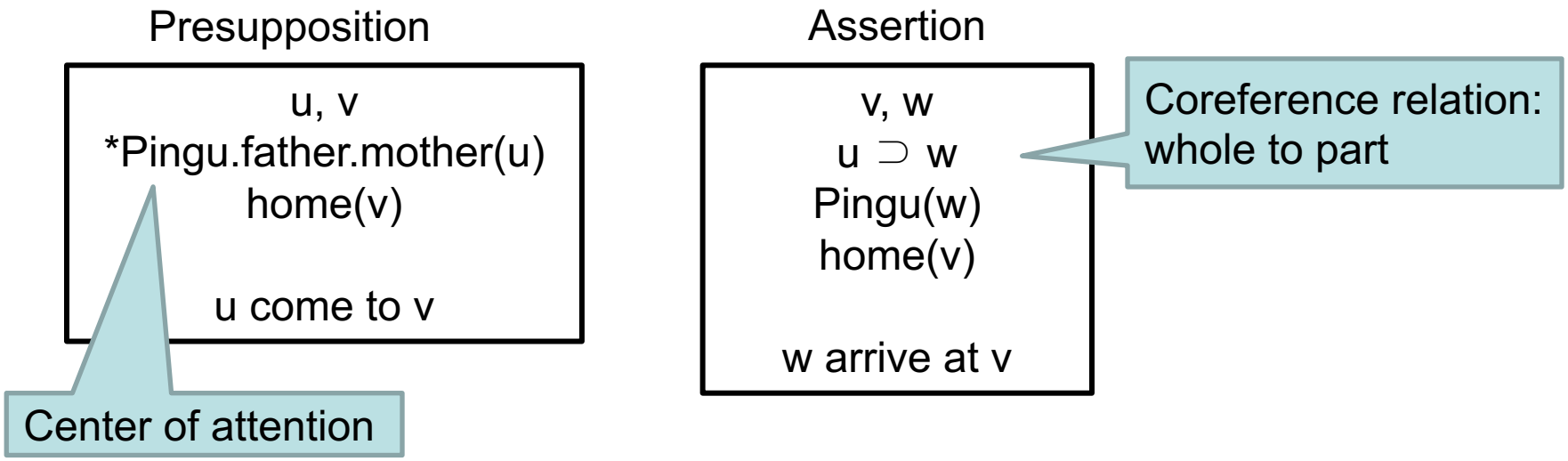
Zero anaphora applies to non-focus arguments only.

Putting all these in RRG representations.

- Discourse representation structures
- Logical structures
- Linking algorithm
- Constructional schema

Discourse representation structures
(A19)

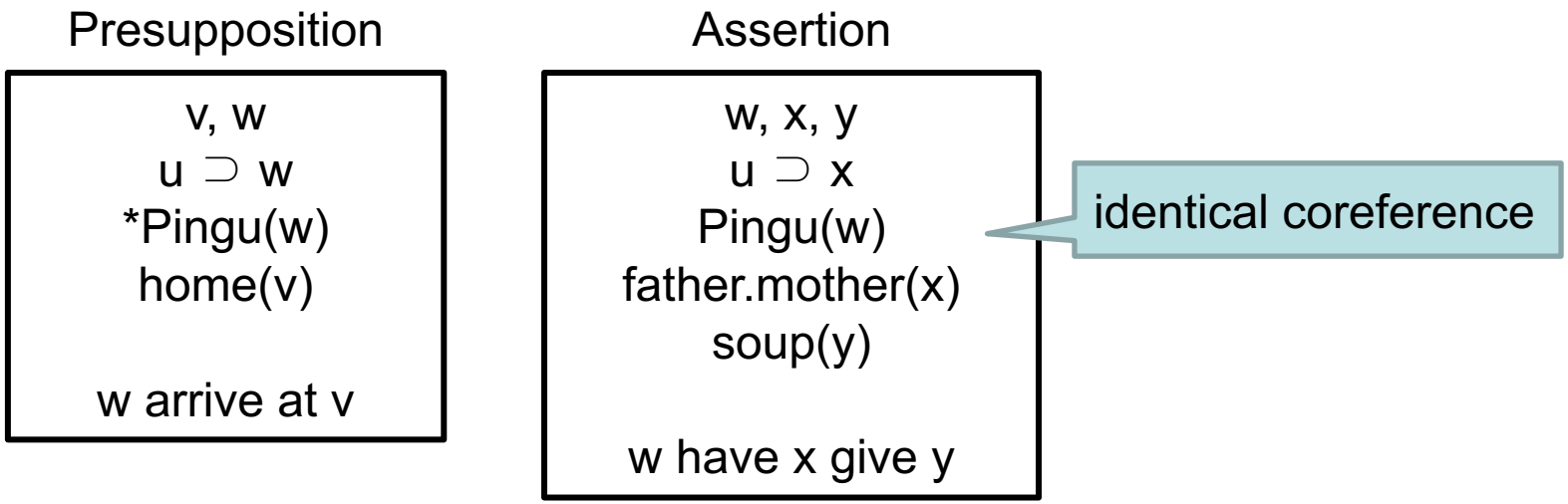
22. sannin-wa [SS] isshoni uchi-ni kaettekita. [Pingu.father.mother, home]
    3.CL-TOP together home-to returned
    'The three (Pingu family) came home together.'
23. pinguu-wa [R] uchi-ni tsuite kara, [Pingu.father.mother/Pingu, home]
    Pingu-TOP home-to arrive after
    'After Pingu arrived home'



(A19)

23. pinguu-wa [R] uchi-ni tsuite kara, [Pingu.father.mother/Pingu, home]
    Pingu-TOP home-to arrive after
    'After Pingu arrived home'

24. (Pingu) [SS] oya-ni …suupu-o nomasetemorai, [Pingu, parents, soup]
    parents-by soup-ACC have.drink
    '(Pingu) had the parents give (him) …soup'



(A19)

24. (Pingu) [SS] oya-ni …suupu-o nomasetemorai, [Pingu, parents, soup]
parents-by soup-ACC have.drink
‘(Pingu) had the parents give (him) …soup’

25. (Pingu) [C] karada-o atatameta. [Pingu, body]
body-ACC warmed.up
‘(Pingu) warmed up (his) body.’

Presupposition	Assertion
w, x, y u ⊃ x *Pingu(w) father.mother(x) soup(y) w have x give y	w, z Pingu(w) body(z) w warm up z

Logical structure

A topic must be represented in the logical structure.

(1) A topic sentence has distinct semantic properties (Shibatani 1990).

hi-ga noboru → perceptual judgment "(Look) the sun rises."

BECOME **risen'** (sun)

hi-wa noboru → experiential judgment "(I know) the sun rises."

The topic modifies the whole clause.

topic' (sun, [BECOME **risen'** (sun)])

(2) A non-argument topic is possible.

sakana-wa tai-ga ii

fish-TOP sea.beram-NOM good

'As for fish, sea bream is good,'

topic' (fish, [**be'** (sea bream, [**good'**]))

The linking algorithm: semantics → syntax

1. Construct the semantic representation of the sentence, based on the logical structure of the predicator.
   If the entity to be predicated represents presupposition and are not to represent the actual focus, add (topic' (x, […]) for the entity.
   (Marked construction) if the entity is to be given as a focus despite the actual non-focus, do not use the topic construction.

2. Determine the actor and undergoer assignments, following the actor-undergoer hierarchy.

For the marked assignment, see Shimojo (2011).

The linking algorithm: semantics → syntax (cont.)

3. Determine the morphosyntactic coding of the arguments.
   a. Select the privileged syntactic argument, based on the privileged syntactic argument selection hierarchy and principles.
   b. Assign the arguments the appropriate case markers and/or postpositions.
      1. Determine the argument form for non-focus, following the saliency hierarchy.
         (Marked assignment) if an argument representing non-focus is to be given as a focus, use an overt form.
      2. Assign *wa* for an overt topic and appropriate case markers for all remaining arguments, based on the case assignment rules for accusative constructions. If the argument requires absolute (i.e. non-contrastive) specification, assign no marking.
      3. (Marked assignment) if an argument in actual focus needs to be defocused, assign *wa*.

The linking algorithm: semantics → syntax (cont.)

4. Select the syntactic template(s) for the sentence, following the syntactic template selection principle (and language-specific qualifications).
   a. If an argument has no syntactic instantiation, use the syntactic template without the corresponding RP node.
   b. Use the LDP for a *wa*-marked element, but place it in a RP if it is in the actual focus domain. Use the PrCS for a *ga* or *wa*-marked narrow-focus argument.
   c. If the referent of an argument is to be defocused, use a PoCS if it is in the actual focus domain or a RDP if it is outside the focus domain.
5. Assign arguments to positions in the syntactic representation of the sentence. If there is no syntactic position to assign the argument(s) to, link them directly with the corresponding referents in the presupposition discourse representation structure.

The selection of a PSA (and possibly the predicate) is influenced by the coherence ranking.

continue > retain > smooth-shift > rough-shift

greater coherence ←

(A19: 24)

Presupposition	Assertion
v, w u ⊃ w *Pingu(w) home(v) w arrive at v	w, x, y u ⊃ x Pingu(w) father.mother(x) soup(y) w have x give y

'Pingu' as PSA → Smooth Shift (greater coherence)

'father.mother' or 'soup' as PSA → Rough Shift.

Further addition to Step 1 of the the algorithm.

The linking algorithm: semantics → syntax

1. Construct the semantic representation of the sentence, based on the logical structure of the predicator.
   If discourse coherence is intended, use a predicator to provide predication for the most salient entity in the presupposition DRS.
   If the entity to be predicated represents presupposition and are not to represent the actual focus, add (**topic'** (x, […]) for the entity.

The linking algorithm: syntax → semantics

1. Determine the macrorole(s) and other core argument(s) in the clause.
2. Retrieve from the lexicon the logical structure of the predicate in the nucleus of the clause and with respect to it determine the actor and undergoer assignments, following the actor-undergoer hierarchy.
   If there is a topic in the LDP or RDP, link (**topic'** (x, […]) to the matrix logical structure.
   If the clause structure contains no predicate, retrieve the predicate from the presupposition discourse representation structure.
3. Link the arguments determined in step 2 with the arguments determined in step 3 until all core arguments are linked.
   If there is an unlinked argument position(s) in the semantic representation, retrieve the corresponding referent(s) directly from the presupposition discourse representation structures.

How do we capture the general tendency that the PSA of a sentence-initial unit tends to be overtly expressed?

Because this is the case across transition types and construction specific, it is represented by a construction schema, like English conjunction reduction (Van Valin 2005: 230).

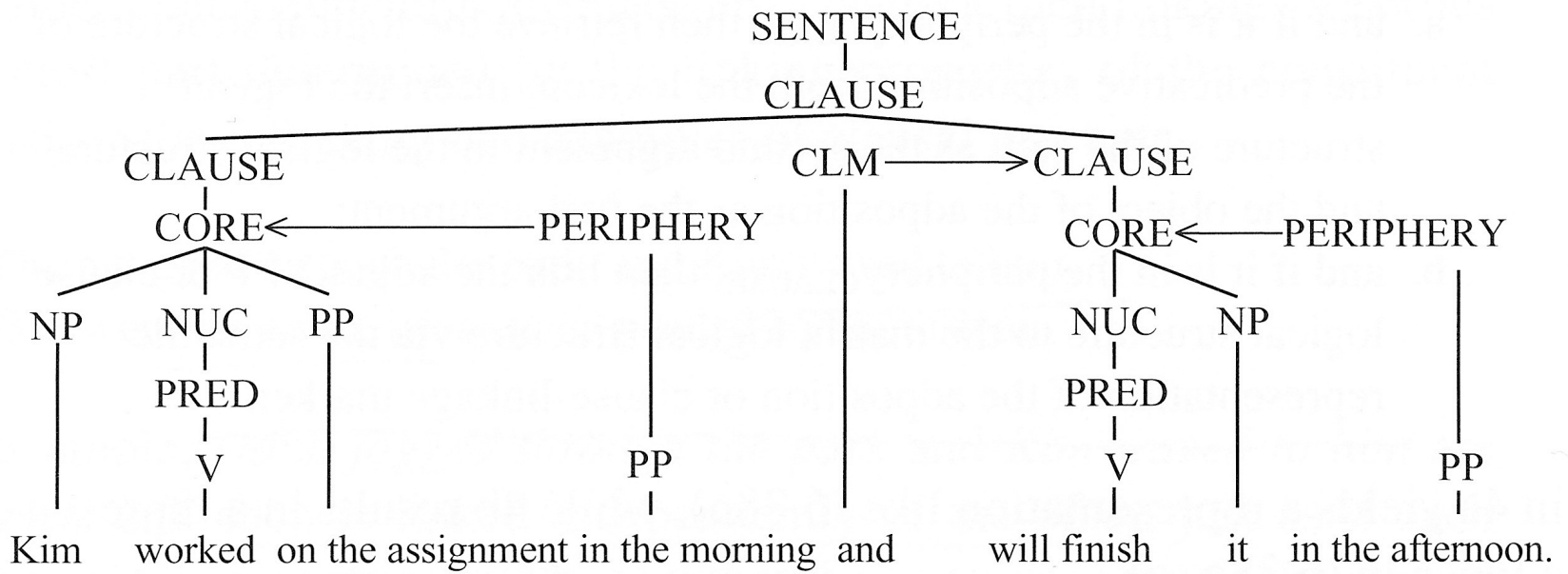


Figure 7.1 *Constituent projection for (7.7)*

Construction schema for Japanese 'topic omission' (simplified)

SYNTAX:

Juncture: clausal

Nexus: cosubordination

Construction type: conjunction

(default) $[_{CL}\ [_{CORE}\ NP_i \dots\ [_{NUC} \dots]]]_1$, (CLM) $[_{CL}\ [_{CORE}\ \_\_i\ \dots\ [_{NUC} \dots]]]_2$,
(CLM) $[_{CL}\ [_{CORE}\ \_\_i \dots]]]_n$

PSA: Clause 1: variable syntactic controller = pragmatic controller

Clause 1+n: variable syntactic pivot = pragmatic pivot

Linking: controlled argument in clause 1+n = pragmatic pivot

SEMANTICS:

Sequence of events sharing a common primary topical participant

PRAGMATICS:

Illocutionary force: shared across all conjuncts

Focus structure: predicate focus in all conjuncts

L2 Japanese-specific characteristics

L2-1	CON-SS	SI	CON-SS	~SI	CON-DS	SI	CON-DS	~SI
Zero	27	0.54	25	0.86	2	0.40	0	
WA	23	0.46	4	0.14	3	0.60	0	
Total	50	1.00	29	1.00	5	1.00	0	
L2-2	**CON-SS**	**SI**	**CON-SS**	**~SI**	**CON-DS**	**SI**	**CON-DS**	**~SI**
Zero	13	0.35	50	0.93	2	0.25	1	0.17
WA	24	0.65	4	0.07	6	0.75	5	0.83
Total	37	1.00	54	1.00	8	1.00	6	1.00
L2-3	**CON-SS**	**SI**	**CON-SS**	**~SI**	**CON-DS**	**SI**	**CON-DS**	**~SI**
Zero	42	0.49	123	0.96	1	0.06	2	0.14
WA	44	0.51	5	0.04	17	0.94	12	0.86
Total	86	1.00	128	1.00	18	1.00	14	1.00
L1	**CON-SS**	**SI**	**CON-SS**	**~SI**	**CON-DS**	**SI**	**CON-DS**	**~SI**
Zero	85	0.38	406	0.93	8	0.12	16	0.38
WA	139	0.62	32	0.07	60	0.88	26	0.62
Total	224	1.00	438	1.00	68	1.00	42	1.00

L2 speakers' inconsistency for a continuing PSA across sentences.
The constructional schema (previous slide) is not yet in place in their grammar; thus, the PSA of a sentence-initial unit is frequently omitted (per linking algorithm step 3).

Acquisition of topic forms in L2 Japanese (L1 English speakers)

and RRG representations

(1) Learn *WA* as the general topic/subject marker (typically the first construction taught in class). Early acquisition and overuse of *WA* (Nakahama 2009, Yagi 1999).

Acquisition of linking algorithm relevant to a topic sentence, to use an overt topic for ANY entity to be predicated that has coreference relations in the DRS.

(2) Learn omission of arguments, for PSAs and non-PSAs.

Adjustment of the linking algorithm to omit salient arguments, including PSAs of sentence-initial units (overuse of zero anaphora).

(3) Use *WA* for the PSA of a sentence-initial clause.

Acquisition of the constructional schema for “topic omission”.

Topic in Korean: mismatch with Japanese

J: toire-wa/*ga doko-ni arimasu-ka?
restroom-TOP/*NOM where-LOC exist-Q
'Where is the restroom?'

WA for 'the restroom' (definite)

K: hwacangsil-*un/i eti-ey iss-eyo?
restroom-*TOP/NOM where-LOC exist-Q

Korean *KA/I* (nominative) may mark focus or non-focus.
(N)UN (topic) for marked contrastiveness.

Discourse data

(1) Bible translations

Marking of matrix subjects in the Gospel According to Mark 1-3

	WA [Japanese topic]	*GA* [Japanese nominative]
(N)UN [Korean topic]	49 (match)	1 (mismatch)
KA/I [Korean nominative]	25 (mismatch)	18 (match)

When is *(N)UN* used in Korean?

Mark 2:23 (Japanese)

aru ansokubi-ni iesu-ga mugibatake-o tootteikareru-to deshitati-wa
one Sabbath-on Jesus-NOM grainfields-ACC passing.through-when disciples-TOP
aruki-nagara mugi-no ho-o tsumihajimeta.
walk-while grain-LK heads-ACC began.picking
'One Sabbath day, when Jesus was passing through the grainfields,
the disciples (TOP) began to pick the heads of grain while walking.'

Mark 2:23 (Korean)

enu ansikil-ey yeyswunim-i milpath sailo cinakasil ttay ceycatul-i hamkkey
one Sabbath-on Jesus-NOM field in pass by when disciples-NOM together
ka-myense milisak-ul calla mekessta.
go-while grain.head-ACC cut-and ate
'One Sabbath day, when Jesus was passing through the grainfields,
his disciples (NOM) walked with him, picking up and eating some heads of grain.'

"Disciples" is discourse-old (most recent reference in v.18).
V.23 starts a new episode "one Sabbath day…"

WA represents hearer-givenness.

The Bible data

Japanese	*GA*	*WA*	Total
Hearer-new, discourse-new	15 (.94)	1 (.06)	16 (1.00)
Hearer-old, discourse-new	0	12 (1.00)	12 (1.00)
Hearer-old, discourse-old	4 (.06)	61 (.94)	65 (1.00)
Korean	*KA/I*	*(N)UN*	Total
Hearer-new, episode-new	14 (.88)	2 (.13)	16 (1.00)
Hearer-old, episode-new	28 (.93)	2 (.07)	30 (1.00)
Hearer-old, **episode-old**	2 (.04)	45 (.96)	47 (1.00)

(N)UN represents episode-givenness.

(2) Picture-based narratives (see slide #10)

Total counts of the subject forms (10 speakers each)

	Japanese		Korean	
Topic	120	(.69)	86	(.53)
Nominative	19	(.11)	63	(.39)
Zero anaphora	35	(.20)	13	(.08)
Zero particle	1	(.01)	1	(.01)
Total	175	(1.00)	163	(1.00)

WA is more frequently used than *NUN*.

KA/I (Korean nom.) is more frequently used than *GA* (Japanese nom.).

Total counts after initial introduction of referents (pictures 2-12)

	Japanese	Korean
Topic	76	50
Nominative	3	46

Total number of topic and nominative NPs by speaker (JPN)

Table 5 Total number of matrix NP forms used by each speaker in the Japanese narratives

Speaker	A	B	C	D	E	F	G	H	I	J	Total
Nom	3 (0.17)	2 (0.15)	3 (0.14)	3 (0.13)	2 (0.13)	3 (0.12)	1 (0.08)	1 (0.08)	1 (0.06)	0	19 (0.11)
Top	12 (0.67)	8 (0.62)	17 (0.77)	11 (0.48)	11 (0.73)	21 (0.81)	9 (0.69)	6 (0.50)	11 (0.69)	14 (0.82)	120 (0.69)
Zero a	3 (0.17)	3 (0.23)	2 (0.09)	9 (0.39)	2 (0.13)	2 (0.08)	3 (0.23)	5 (0.42)	3 (0.19)	3 (0.18)	35 (0.20)
Zero p									1 (0.06)		1 (0.01)
Total	18 (1.00)	13 (1.00)	22 (1.00)	23 (1.00)	15 (1.00)	26 (1.00)	13 (1.00)	12 (1.00)	16 (1.00)	17 (1.00)	175 (1.00)

- The frequency of *WA* and *GA* is relatively uniform across the speakers; *WA* always outnumbers *GA*.

Total number of topic and nominative NPs by speaker (KOR)

Table 6 Total number of matrix NP forms used by each speaker in the Korean narratives

Speaker	A	B	C	D	E	F	G	H	I	J	Total
Nom	12 (0.86)	12 (0.80)	7 (0.41)	7 (0.41)	6 (0.40)	6 (0.33)	5 (0.29)	3 (0.19)	2 (0.17)	3 (0.14)	63 (0.39)
Top	2 (0.14)	2 (0.13)	10 (0.59)	10 (0.59)	6 (0.40)	12 (0.67)	12 (0.71)	9 (0.56)	6 (0.50)	17 (0.77)	86 (0.53)
Zero a		1 (0.07)			2 (0.13)			4 (0.25)	4 (0.33)	2 (0.09)	13 (0.08)
Zero p					1 (0.07)						1 (0.01)
Total	14 (1.00)	15 (1.00)	17 (1.00)	17 (1.00)	15 (1.00)	18 (1.00)	17 (1.00)	16 (1.00)	12 (1.00)	22 (1.00)	163 (1.00)

- Speakers switched from *(N)UN* to *KA/I* to mark the same referents across episodic boundaries (often shown by paragraph breaks).

- Considerable individual variation: some used more *(N)UN* and others used more *KA/I*, depending on how many episodic units are used in the story.

Table 6 Total number of matrix NP forms used by each speaker in the Korean narratives

Speaker	A	B	C	D	E	F	G	H	I	J	Total
Nom	12 (0.86)	12 (0.80)	7 (0.41)	7 (0.41)	6 (0.40)	6 (0.33)	5 (0.29)	3 (0.19)	2 (0.17)	3 (0.14)	63 (0.39)
Top	2 (0.14)	2 (0.13)	10 (0.59)	10 (0.59)	6 (0.40)	12 (0.67)	12 (0.71)	9 (0.56)	6 (0.50)	17 (0.77)	86 (0.53)
Zero a		1 (0.07)			2 (0.13)			4 (0.25)	4 (0.33)	2 (0.09)	13 (0.08)
Zero p					1 (0.07)						1 (0.01)
Total	14 (1.00)	15 (1.00)	17 (1.00)	17 (1.00)	15 (1.00)	18 (1.00)	17 (1.00)	16 (1.00)	12 (1.00)	22 (1.00)	163 (1.00)

More ← Episodic division → Less

Typical episodic structure (speakers C, D, E, F, G)
Episode 1: Husband's meeting a new woman (frame 1-5)
Episode 2: Couple's falling out (frame 6-9)
Episode 3: The man's new life (frame 10-12)

More episodes (speakers A, B)
Each picture was described as a separate episode.

Less episodes (speakers H, I, J)
The whole story as one episode.
Topic outnumbers nominative, similar to Japanese.

Individual variation is expected depending on how the story is presented.

A story example and typical episodic structure used in Korean

E1
1. A couple is sitting on the couch in the living room. The husband is smoking and the wife is reading.
2. The husband looks in the mirror and the wife is reading.
3. He goes to a night club.
4. He meets a young woman and dances with her.
5. And they have dinner at a restaurant.

E2
6. At home, he calls the woman. The wife walks in.
7. He confesses his affair to his wife.
8. She is crying on the bed.
9. He packs his bag to leave the house.

E3
10. In the new place, the husband and the young woman are bored sitting on the couch.
11. She looks in the mirror and the man is sitting.
12. She goes to the night club.

Discourse representation structures

1. A couple is sitting on the couch in the living room.

Assertion

E1, o, p, q
husband.wife(o)
couch(p)
living room(q)

o sit on p in q

The husband is smoking and the wife is reading.

Presupposition

E1, o, p, q
*husband.wife(o)
couch(p)
living room(q)

o sit on p in q

Assertion

E1, r, u
a ⊃ r
a ⊃ u
husband(r)
wife(u)

r smoke
u read

5. And they have dinner at a restaurant.

Presupposition

E1, r, v
*husband(r)
woman(v)

r meet v
r dance with v

Assertion

E1, w, x, y
r ⊂ w
v ⊂ w
husband.woman(w)
dinner(x)
restaurant(y)

w have x at y

6. At home, he calls the woman.

Presupposition

E1, w, x, y
r ⊂ w
v ⊂ w
*husband.woman(w)
dinner(x)
restaurant(y)

w have x at y

Assertion

E2, r, v, z
husband(r)
woman(v)
home(z)

r call v at z

New episode

Revision of the linking algorithm: semantics → syntax

1. Construct the semantic representation of the sentence, based on the logical structure of the predicator. If the entity to be predicated represents presupposition within the current episode or if the entity is in a contrastive relation, add (**topic'** (x, […]) for the entity.
2. Determine the actor and undergoer assignments, following the actor-undergoer hierarchy.
3. Determine the morphosyntactic coding of the arguments.
   a. Select the privileged syntactic argument, based on the privileged syntactic argument selection hierarchy and principles.
   b. Assign the arguments the appropriate case markers and/or postpositions.
      (i) Assign *(n)un* if there is a topic.
      (ii) Assign appropriate case markers for all remaining arguments, based on the case assignment rules for accusative constructions.

*Linking related to zero anaphora is excluded.

Japanese *GA* to re-introduce discourse-old referents afresh (Maynard 1980).

Momotaro “Peach Boy” (English translation)

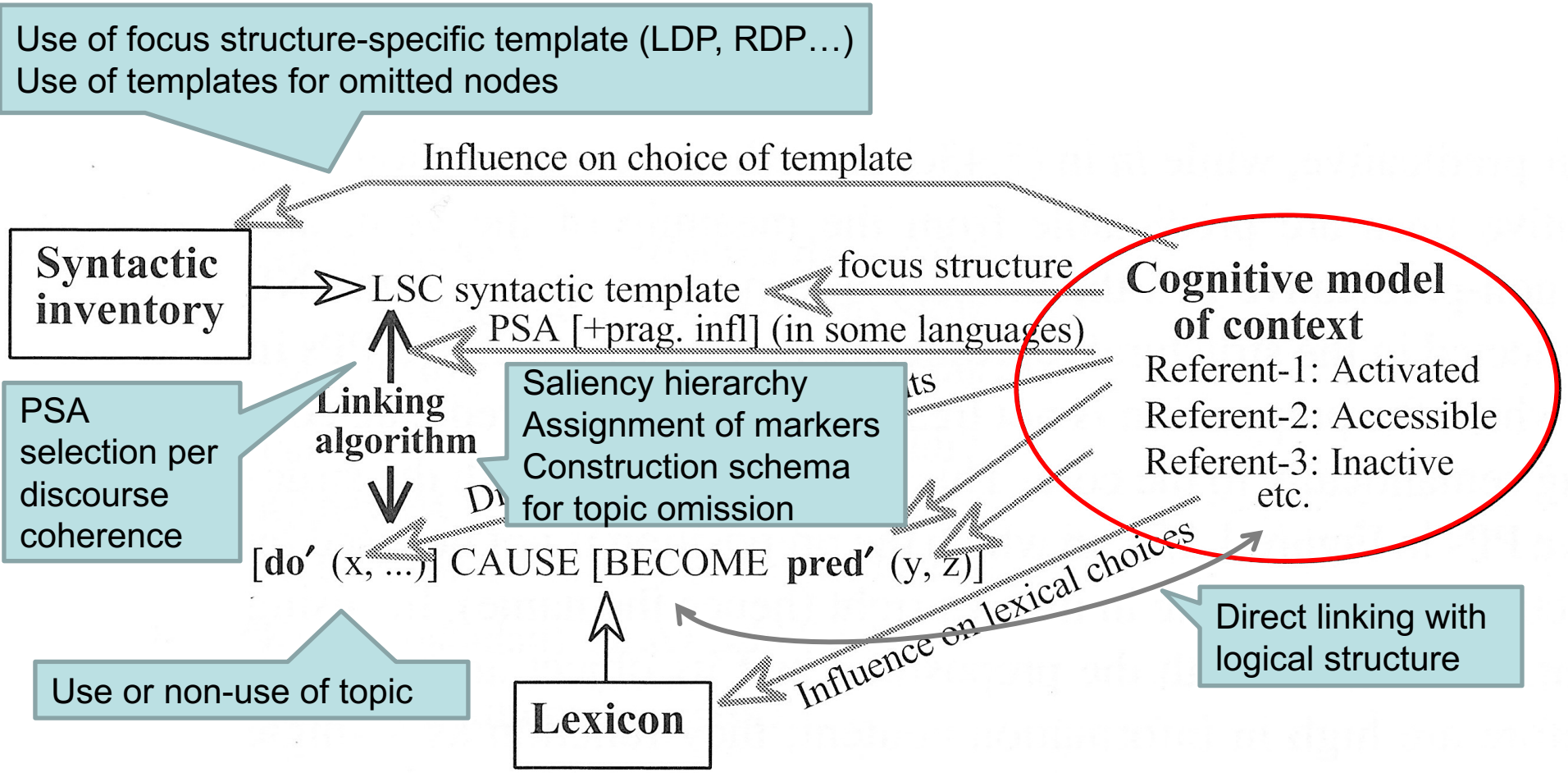
Once upon a time in Japan, there lived in the country an old man and his wife. They were very lonely because they had no children.

One day the old man went into the mountains to cut firewood and his wife went to the river to wash clothes…

GA may be used here.

How is this different from Korean *KA/I* marking for a new episode?

- *GA* achieves a more dramatic effect than *KA/I*.
- *GA* is a focus marker; the use for presupposed referents is a direct conflict with the focus structure.
- *KA/I* is not stored with specific focus structures and the use for presupposition only indicate an episodic update.



Cross-linguistic variation in the interaction with pragmatics

Special thanks to Miho Fujiwara, Mitsuko Yamura-Takei, and Etsuko Yoshida, for sharing the animation-based Japanese narrative data, and EunHee Lee for the Korean data.

Brown, Gillian; George Yule. 1983. Discourse Analysis. Cambridge: Cambridge University Press.

Chafe, Wallace. 1994. Discourse, Consciousness, and Time. Chicago: University of Chicago Press.

Fry, John. 2003. Ellipsis and WA-marking in Japanese Conversation. New York & London: Routledge.

Grosz, Barbara, Aravind K. Joshi, and Scott Weinstein. 1995. Centering: a framework for modeling the local coherence of discourse. Computational Linguistics 21(2), 203-225.

Grosz, Barbara, and Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. Computational Linguistics 12, 175-204.

Gundel, Jeanette, Nancy Headberg, and Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. Language 69, 274-307.

Hinds, John, and Wako Hinds. 1979. Participant Identification in Japanese Narrative Discourse. In Explorations in Linguistics: Papers in Honor of Kazuko Inoue, G. Bedell et al. (eds.), Tokyo: Kaitakusha, 201-12.

Kuno, Susumu. 1973. Nihon bunpoo kenkyuu (Studies in Japanese grammar). Tokyo: Taisyuukan.

Lee, EunHee, and Mitsuaki Shimojo. 2016. Mismatch of topic between Japanese and Korean. Journal of East Asian Linguistics 25: 81-112.

Maynard, Senko K. 1980. Discourse functions of the Japanese theme marker - wa. Ph.D. dissertation, Northwestern University.

Nakahama, Yuko. 2009. Cross-linguistic influence on referent introduction and tracking in Japanese as a second language. Modern Language Journal 93, 241-260.

Prince, Ellen F. 1992. The ZPG letter: Subjects, definiteness, and information status. In W. Mann and S. Thompson (eds.), Discourse description: Diverse linguistic analyses of a fund-raising text, 295-325. Amsterdam: Benjamins.

Shibatani, Masayoshi. 1990. The Languages of Japan. Cambridge University Press.

Shimojo, Mitsuaki. 2005. Argument Encoding in Japanese Conversation. Hampshire and New York: Palgrave Macmillan.

Shimojo, Mitsuaki. 2011. The Left Periphery and Focus Structure in Japanese. In New Perspectives in Role and Reference Grammar, Wataru Nakamura (ed.), Newcastle: Cambridge Scholars Publishing, 266-293.

Suzuki, Satoko. 1995. The functions of topic-encoding zero-marked phrases: a study of the interaction among topic-encoding expressions in Japanese. Journal of pragmatics 23, 607-626.

Van Valin, Robert D. Jr.; Randy J LaPolla. 1997. Syntax: Structure, Meaning & Function. Cambridge: Cambridge University Press.

Van Valin, Robert D. Jr. 2005. Exploring the syntax-semantics interface. Cambridge University Press.

Van Valin, Robert D. Jr. 2014. On the place of information structure in a grammar. Comunicación, Cognición, Cibernétic@. Actas del XXXI Congreso de AESLA, 86-106.

Walker, Marilyn A., Masayo Iida, and Sharon Cote. 1994. Japanese discourse and the process of centering. Computational Linguistics 20(2), 193-231.

Walker, Marilyn A., Aravind K. Joshi, and Ellen F. Prince. 1998. Centering in naturally occurring discourse: An overview. In M. A. Walker, A. K. Joshi, and E. F. Prince, editors, Centering Theory in Discourse. Oxford University Press, Oxford, chapter 1, pages 1–28.

Watanabe, Yasuko. 1989. The Function of WA and GA in Japanese Discourse. Ph.D. dissertation, University of Oregon.

Yagi, Kimiko. 1999. Universal topic-prominence in interlanguage: from the perspective of SLA research on ”wa” and “ga”. Acquisition of Japanese as a Second Language 2, 57-67.