DIFFERENTIAL CASE MARKING - II

OT Syntax and Typology

Summer School, Düsseldorf
Judith Aissen

THE GENERALIZATION UNDERLYING DOM

The higher in prominence a direct object, the more likely it is to be overtly case marked.

Prominence assessed on two scales:

Animacy:

Human > Animate > Inanimate

Definiteness:

Pronoun > PN > Definite > Indefinite Specific > NonSpecific

THE INTUTION BEHIND THE ANALYSIS

High rank on these scales is unmarked (frequent) for subjects, but marked (infrequent) for objects. There is thus a bias to interpret high-ranked nominals as subjects. If they are in fact objects, DOM counteracts this bias.

THE ANALYSIS

Since we already have constraints to characterize marked and unmarked objects – constraints derived by $\mathcal{HA}$ -- they are used here…

The Analysis

Key Constraints for DOM

• Constraint subhierarchies based on $\mathcal{HA}$

*Oj/Pro & Case » … » *Oj/NSpec & Case

*Oj/Hum & Case » … » *Oj/Inan & Case

Penalize the absence of morphological case.

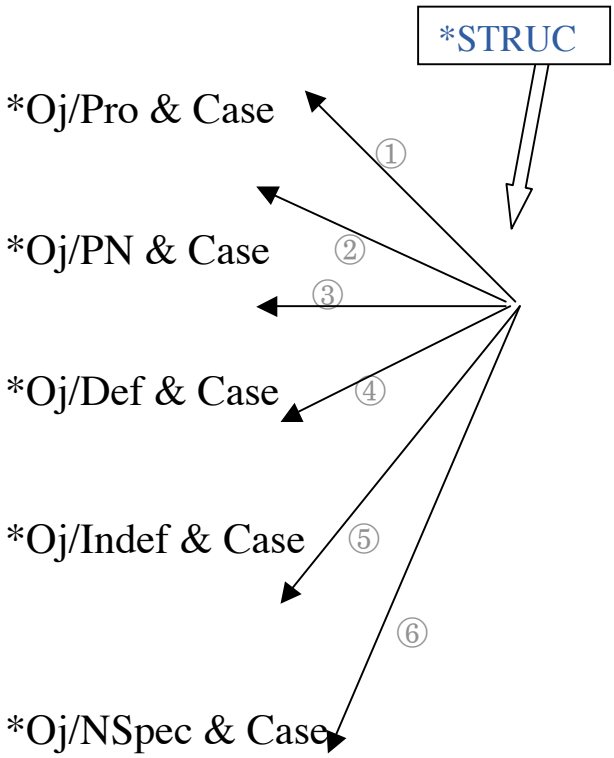
• *Struc

Penalize the presence of morphological case

The constraint subhierarchies are universal. What is language-particular is the point at which *Struc interpolates. *Struc functions to mark the point in the hierarchy where case is 'turned off'.

> The role of Input-Output Faithfulness: No role here, since this account assumes that case is not specified in inputs.
>
> For accounts which assume case is present in inputs, and which make crucial use of IO Faithfulness *and of* $\mathcal{HA}$, see (Stiebels 2000; Stiebels 2000; Ortmann to appear).

The typology of definiteness-sensitive DOM



① Kalkatungu: no objects case-marked.

② Catalan: only pronouns case-marked

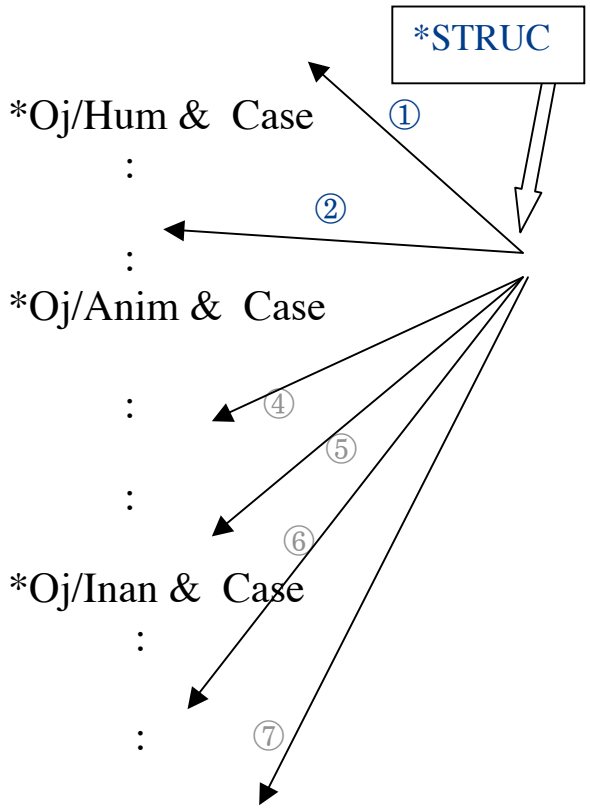
③ Pitjantjatjara: only pronouns and PNs case-marked.

④ Hebrew: all and only definites case-marked.

⑤ Turkish: all and only indefinite specifics case-marked.

⑥ Written Japanese, Dhalandji: all objects case-marked

The Typology of Animacy-sensitive DOM



① Kalkatungu: no objects case-marked.

② Yiddish: only some human objects case-marked.

④ Ritharngu: all human objects and some animates case-marked.

⑤ Dhargari: all animate objects case-marked. Sinhalese, all animate objects optionally case-marked.

⑥ Bayungo: all animate and some inanimates case-marked.

⑦Written Japanese, Dhalandji: all objects case-marked.

• The inventory of morphological cases in a language and the distribution of those cases are *emergent* properties, properties which are a consequence of the language's constraint ranking.

Typology of Object Marking Systems Predicted

No accusative marker:

*Struc » *Oj/Human & Case, *Oj/Pro & Case » …

All objects case-marked (‘pure’ accusative language):

…» *Oj/Inan & Case, *Oj/NSpec & Case » *Structure »

Differential object marking:

*Struc interpolated among the *Oj/X & Case constraints.

ℋA ACCOUNT OF DOM PREDICTS THE EXISTENCE OF DIFFERENTIAL SUBJECT MARKING (DSM)

The functional motivation for DOM

> To facilitate the distinguishing of subject and object. The properties which increase the likelihood of overt case marking for objects are exactly those most frequently associated with subjects.

If this is right, then Differential Subject Marking (DSM) should exist

> DSM should be found with subjects of *low* prominence (indefinites, inanimates, 3rd persons, non-pronouns). For these are the properties most frequently associated with objects.

The formal system predicts DSM:

> Every Object-oriented subhierarchy is paired with a Subject-oriented one running in the opposite direction.

*Su/Inan & Case » *Su/Anim & Case » *Su/Hum & Case

*Su/NSpec & Case » *Su/Indef & Case … *Su/Pro & Case

Penalize the absence of morphologic case

*Struc

Penalize morphological Case

Predicts:

- Languages in which all subjects are case-marked.
- Languages in which no subjects are case-marked.
- Languages in which low prominence subjects are case-marked, but not high ones (*Differential subject marking)*

EXAMPLES OF DIFFERENTIAL SUBJECT MARKING:

Human > Animate > Inanimate 1st/2nd Pro > 3rd Pro > PN > Definite > Indefinite Specific > Non-Specific

- Dyirbal and Punjabi, in which all transitive subjects are overtly case marked except 1st and 2nd person pronouns.

- Guugu Yimidhirr, in which all lexical NP subjects in transitive clauses are overtly case marked, but personal pronouns are not.

- Fore in which neither personal pronouns *nor* names may be marked in transitive subject function, but inanimates must be. Elements between these two extremes may be.

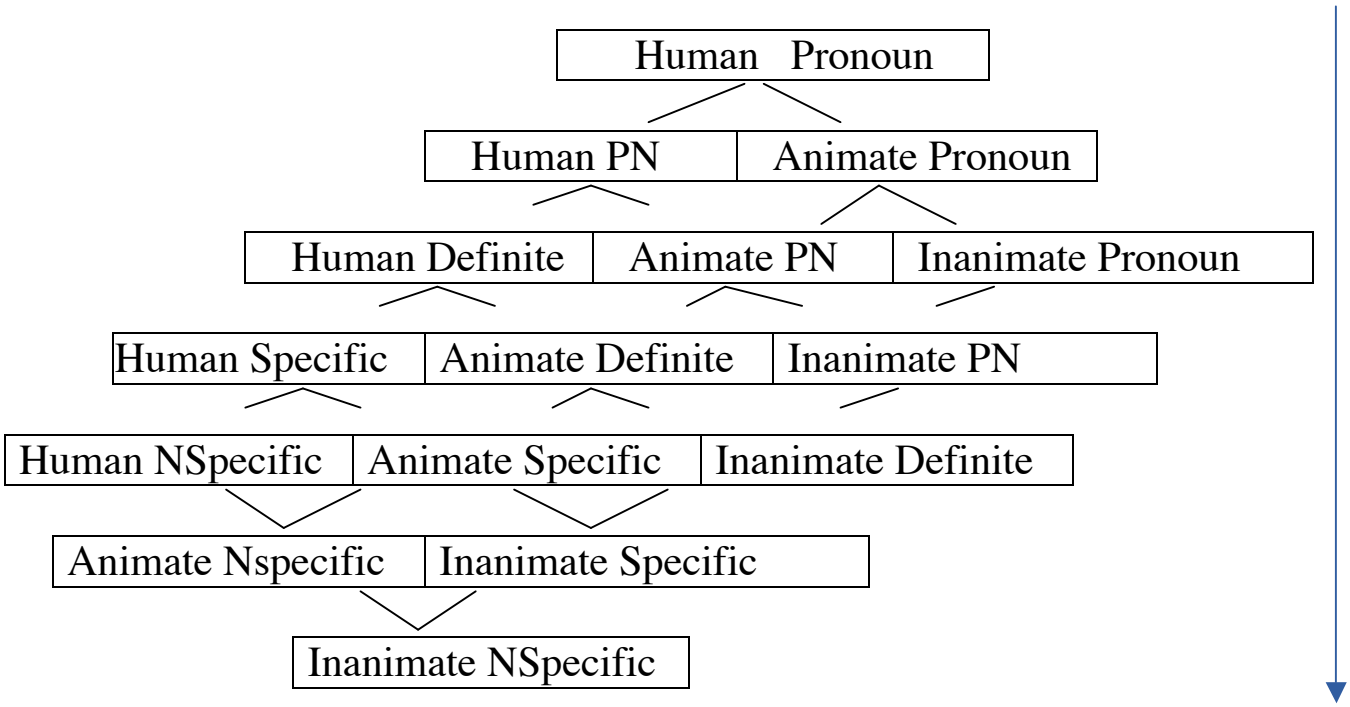
Note: Because these languages show overt case marking only for transitive subjects, the case is called 'ergative'. And because it effects only *some* transitive subjects, the phenomenon has been called 'split ergativity'.

But this 'split ergativity' is just Differential Subject Marking in transitive clauses.

TWO-DIMENSIONAL DOM

In many languages, DOM references both animacy and definiteness, e.g.

I-E/Romance:	Spanish, Romanian
I-E/Germanic:	Colloquial Afrikaans
I-E/Indic:	Hindi, Bengali, Kashmiri
Dravidian:	Tamil, Malayalam, Kannada
Hokan:	Eastern and Northern Pomo
Pama-Nyungan:	Kalaw Lagaw Ya
Japanese/Korean:	Colloquial Japanese, Colloquial Korean

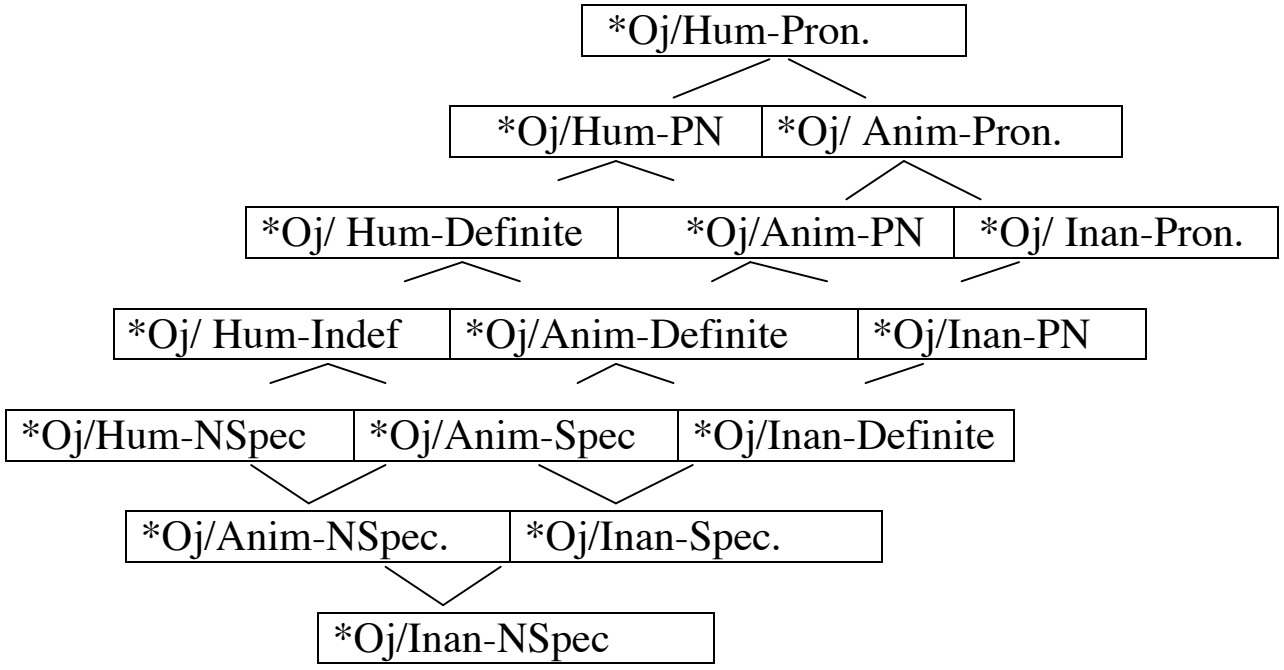


- If □, an object, can be overtly case-marked, then (all else equal) an object more prominent than □ can be overtly case-marked.
- If □, an object, must be overtly case-marked, then (all else equal) an object more prominent than □ must be overtly case-marked.

Modelling two-dimensional DOM systems

A partially ordered constraint set isomorphic to the above lattice is derived by local conjunction of the constraints on object definiteness and object animacy and further conjunction with Case (Case is omitted below).

[Read: **Oj/Hum-Pron.* as [[*Oj/Human $\&_{DP}$ *Oj/Pronoun] & Case]



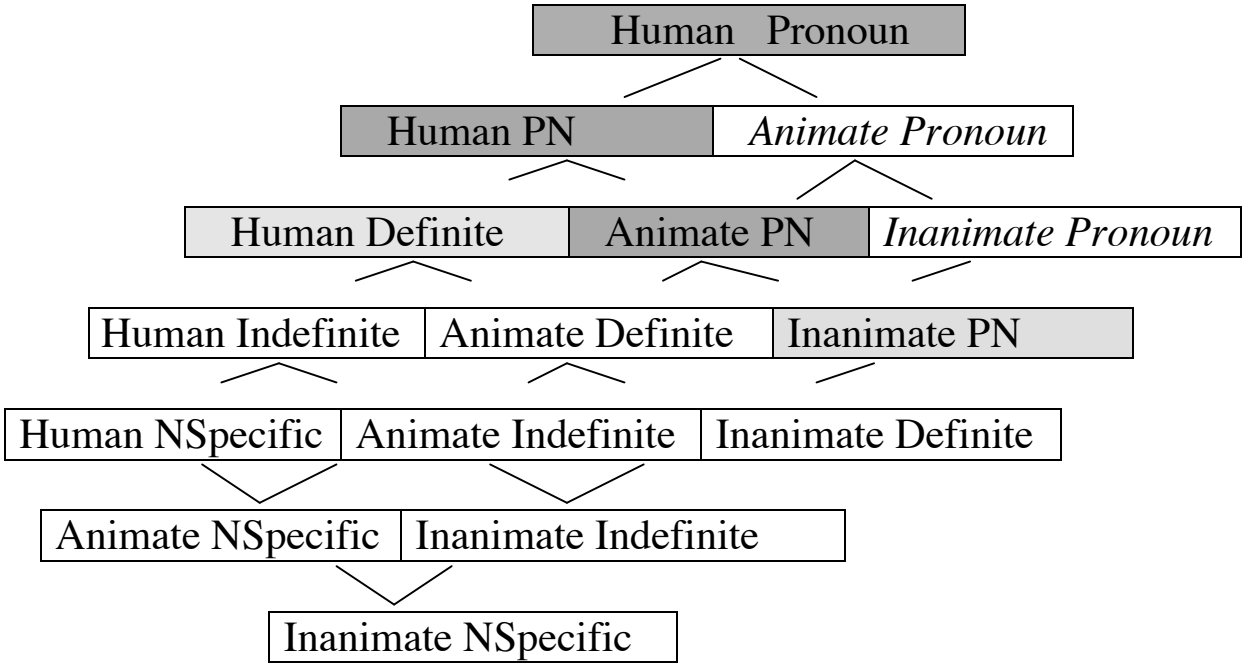
How DOM is realized in particular languages depends on the position of *Struc in the partial ranking represented by this lattice.

VARIABLE DOM

A very common property of two-dimensional DOM systems is that they sort object expressions into three classes:

- Those which require case-marking.
- Those which allow case-marking.
- Those which preclude case-marking.

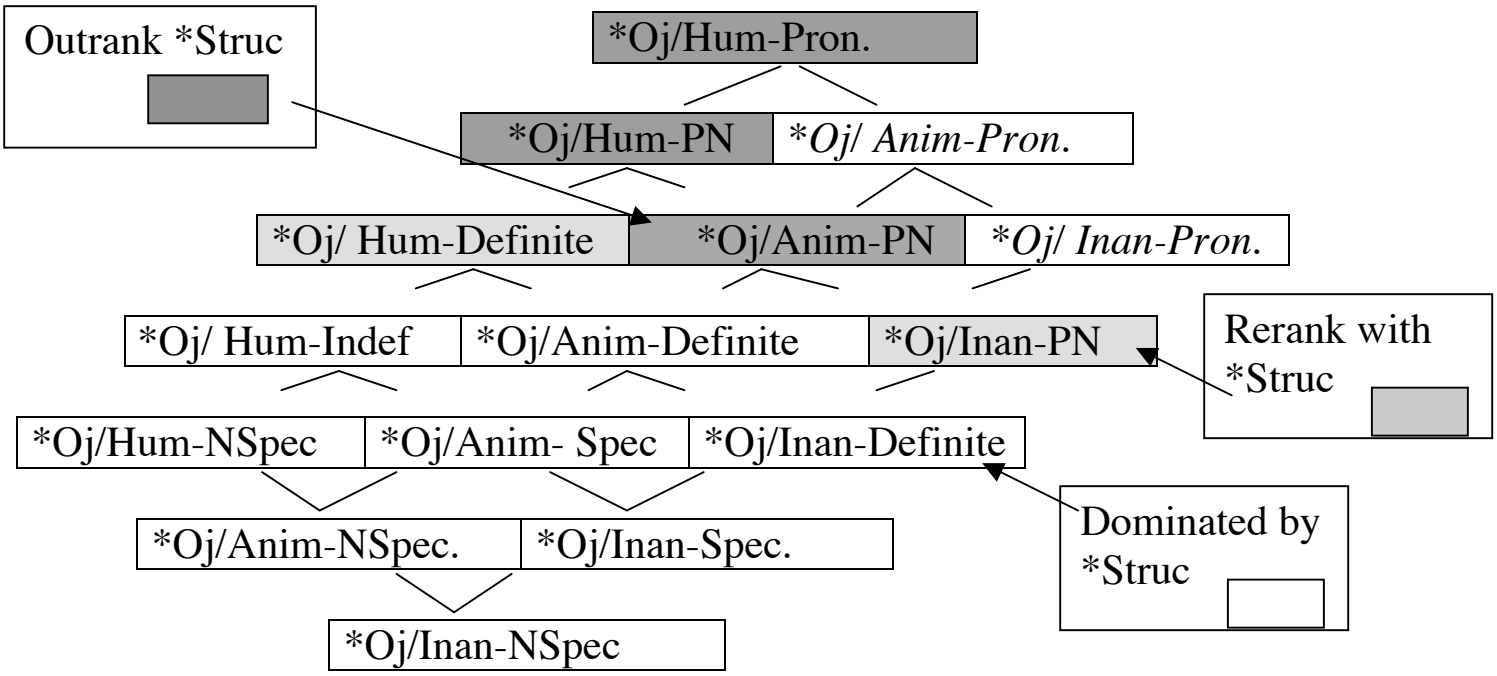
DOM in 12th century Spanish (Cantar de Mío Cid)[1]



[1] Dark cells: obligatory case marking; light cells: optional case marking; white cells: no case marking. In the 12th C. Spanish, I ignore the distribution of DOM on pronouns other than personal pronouns.

DOM in particular languages depends on the position of *Struc in the partial ranking represented by this lattice.

- 'Obligatory' cells outrank, and are relatively distant from, *Struc.
- 'Optional' cells are relatively close to *Struc (and can thus rerank with it).
- 'Impossible' cells are outranked by, and relatively distant from, *Struc.



Since the phenomena is characterized simply by the position of *Struc in the partial ranking of the lattice, it follows (from transitivity of constraint ranking) that if there is an optional DOM zone, it will fall "between" the obligatory and the impossible zones.

The Expansion of DOM to Human-Definite Objects in Spanish
[BASED ON LACA 2000]

	Stage I Prior to 12th C.	Stage II 12th C - 1830	Stage III Contemporary
Hum-Pron	Obligatory	Obligatory	Obligatory
Hum-PN	Obligatory	Obligatory	Obligatory
Hum-Def	*	Optional	Obligatory
Hum-Indef	*	*	Optional

Consider a non-stochastic OT approach to optionality in which optionality results from the fact that the ranking between two constraints (A, B) is not fixed by the grammar (the constraints are "tied") (Müller 1999).

There are evaluations in which A » B as well as ones in which B » A.

Stage 1: … *Struc » *Oj/Hum-Def & Case » …

Stage 2: *Struc «» *Oj/Hum-Def & Case » …

Stage 3: *Oj/Hum-Def » *Struc » …

- This approach to optionality can only describe three stages.
- Change appears to be abrupt, and Stage 2 appears homogeneous.
- However there is historical evidence which shows that the expansion of DOM to Hum-Definite objects proceeded gradually.

THE EXPANSION OF DOM IN THE DOMAIN OF HUMAN-DEFINITE OBJECTS IN SPANISH

[EACH COUNT BASED ON ONE TEXT FROM PERIOD]

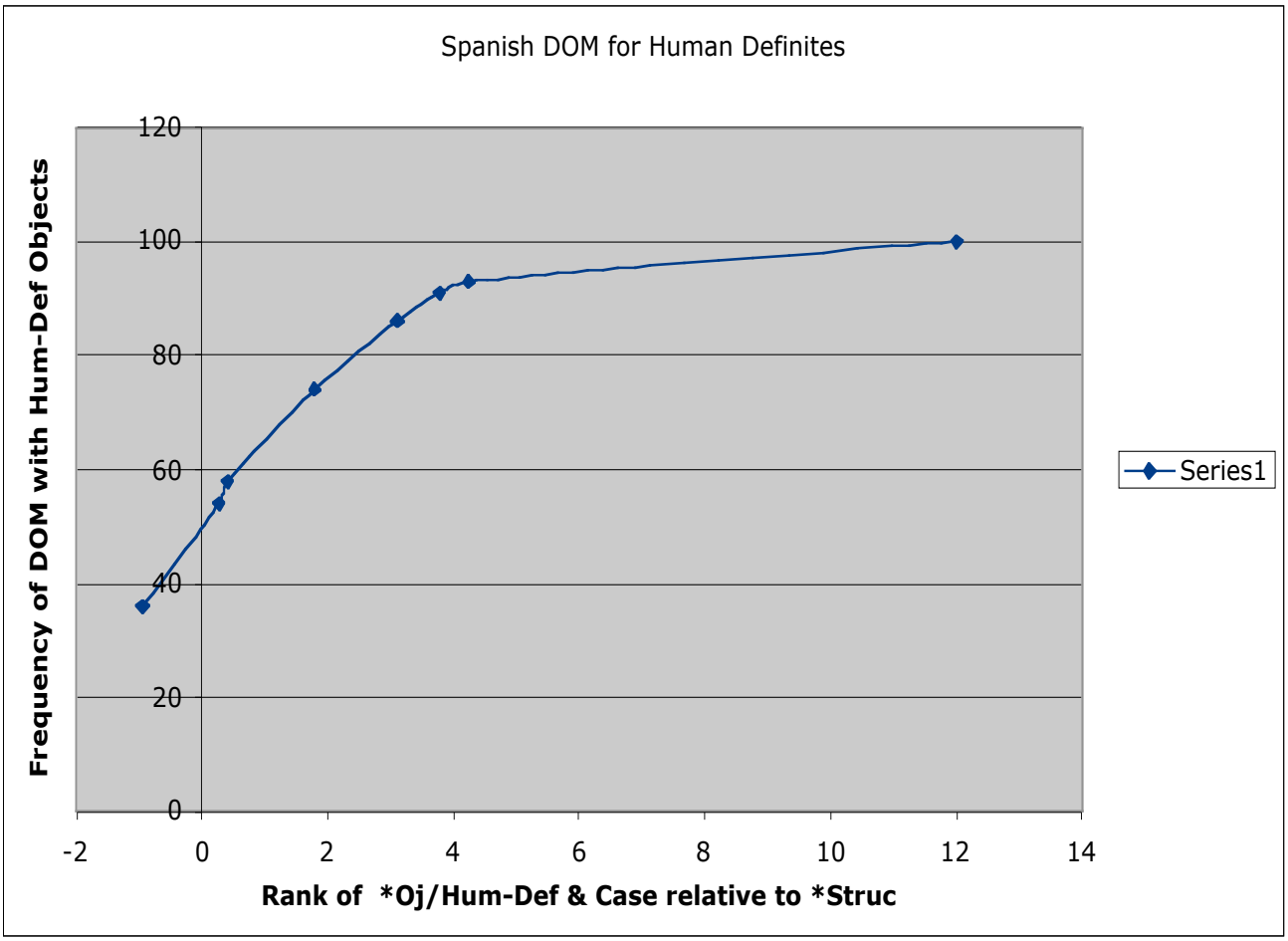
	% DOM with Hum-Def. Oj.
12th C	36%
14th C	54%
15th C	58%
16th C	74%
17th C	86%
18th C	91%
1830	93%
1870	100%

USUAL ASSUMPTION:

Present-day	100%

A Sequence of (Partial) Stochastic OT Grammars which characterizes the gradual extension of Dom to Human-Def Objects

	% DOM with Hum-Def. Oj.	Rank Difference between [*Oj/Def-Hum & Case] and *Struc
12th C	36%	-.953
14th C	54%	.264
15th C	58%	.405
16th C	74%	1.79
17th C	86%	3.107
18th C	91%	3.778
1830	93%	4.24
1870	100%	12.004


Spanish DOM for Human Definites
Frequency of DOM with Hum-Def Objects
120
100
80
60
40
20
0
-2
0
2
4
6
8
10
12
14
Rank of *Oj/Hum-Def & Case relative to *Struc
Series1

MOTIVATION FOR THE EXPANSION OF DOM
(Zeevat and Jäger 2002)

There is functional pressure within the system to extend DOM downwards. For example, once an object marker for definite humans (say) exists, the absence of the marker on definite humans will tend to be interpreted as indicating a subject. The more object-types that can be marked, the stronger the pressure to interpret unmarked NPs as subjects.

More precisely:

Suppose that in the absence of case marking, we have these probabilities:

p(subj | hum def) = 60%

p(obj | hum def) = 40%

But suppose that 36% of human definite objects are case marked. Then the probabilities change:

p(subj | hum def) = 70%

p(obj | hum def) = 30%

To counteract the now stronger bias to interpret (unmarked) human definites as subjects, the frequency of DOM increases within that category, eventually saturating it.

The existence of the marker for definite humans also increases the pressure to mark indefinite humans and definite non-humans.

Conclusion:

The Stochastic Generalization of OT allows the expansion of DOM in the domain of human definite objects to be modelled as a gradual change in the relative ranking of [*Oj/Def-Hum & Case] and *Struc.

Further, there is a well-motivated functional explanation for the gradual increase in distance between the two constraints.

Question:

This may be a reasonable way to characterize the typology of DCM and the historical spread of DCM. But what evidence is there that the full hierarchy of constraints is present in the grammars of individual languages (or individual speakers)?

Take Japanese, for example, as reflected by the written language. Case marking of subjects and objects is obligatory. Could it be plausible that the grammars of Japanese speakers contain the full subhierarchies which determine the typology of DOM and DSM?

Answer:

Yes – it could be! There is evidence in some languages that speakers' grammars do contain more of the structure characterized by the DCM subhierarchies than one might imagine… The evidence again comes from variable DCM.

Languages with variable DOM or DSM extending over several nominal types do not have an unstructured optional zone. Rather, the frequency of DCM in optional zones mirrors the typological distribution of categorical systems:

- the higher in prominence the object, the *more frequently* it is case marked.
- the lower in prominence the subject, the *more frequently* it is case marked.

The same structure which underlies the cross-linguistic typology of DCM systems underlies variable DCM.

PERSIAN

The suffix *–râ* marks some but not all direct objects. Lazard distinguishes three values for definiteness, which correspond to what I am calling DEFINITE, SPECIFIC, and NON-SPECIFIC. Definites are obligatorily suffixed with *-râ*, regardless of animacy, e.g.

Ketâb-râ xândam.
book-ACC I.read
I read the book. (Lazard 1982, 181)

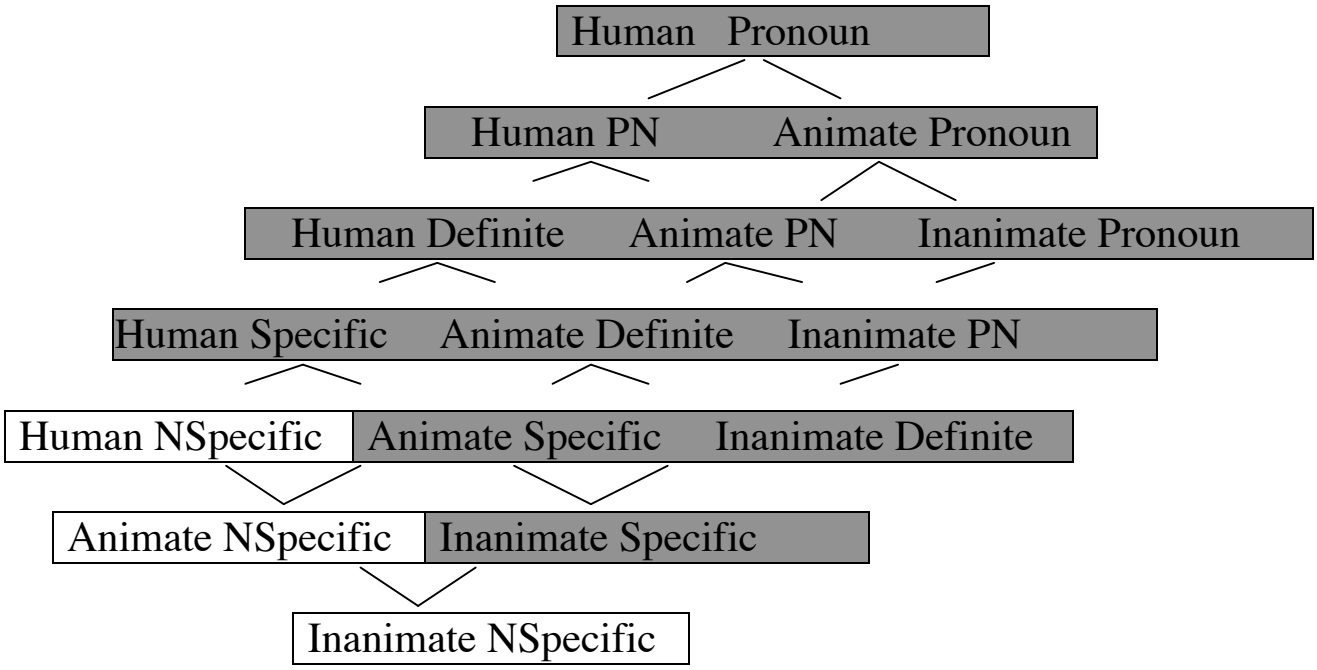
Indefinites are optionally marked, but according to Lazard, specific indefinites are, as a rule, marked. He notes two classes of specific indefinites: those which have a partitive sense, and those with the sense of *a certain*. Both classes require the suffix *–râ*.

Yeki az ân ketâbhâ-râ xândam.
INDEF of DEM books-ACC I.read
I read one of these books. [Lazard, 1982, 183]

(Yek) ketâb-i-râ xând ke…
INDEF book-INDEF-ACC he.read which
He read a certain book which… [Lazard, 1982, 183]

In contrast to both Spanish (12th c. and Modern) and Hindi, this is so irregardless of the animacy of the object. Persian, like Turkish then, requires case-marking for all specific objects.

With respect to other non-specific indefinites, however, Persian appears to be different from Turkish. For the class as a whole, case-marking is optional, but within the class of non-specifics, the distribution of case-marking is determined by animacy. According to Lazard, the probability of case-marking decreases sharply as one moves from human to inanimate. The result is that marking is found generally with humans or animates, but not with inanimates.


Human Pronoun
Human PN
Animate Pronoun
Human Definite
Animate PN
Inanimate Pronoun
Human Specific
Animate Definite
Inanimate PN
Human NSpecific
Animate Specific
Inanimate Definite
Animate NSpecific
Inanimate Specific
Inanimate NSpecific

based on (Fry 2001), which was based on the annotated CallHome Japanese corpus

N1 = subject, N2 = direct object

Following particle?	Animate				Not animate			
	SU		OJ		SU		OJ	
yes	1,642	.65	208	.54	1926	.70	1,117	.47
no	873	.35	178	.46	829	.30	1,253	.53
Total	2,515	1.00	386	1.00	2,755	1.00	2,370	1.00

Particle ellipsis and animacy in CHJ (Fry 2001, 128)

Different rates of particle ellipsis in animate and inanimate subjects is statistically significant.[2]

Following particle?	Proper Noun or personal pronoun				Other			
	SU		OJ		SU		OJ	
yes	918	.63	104	.59	2,650	.70	1,221	.47
no	545	.37	72	.41	1,157	.30	1,359	.53
Total	1,463	1.00	176	1.00	3,807	1.00	2,580	1.00

Particle ellipsis and strongly definite NPs in CHJ ([Fry 2001, 128]

Different rates of particle ellipsis in 'strongly' definite subjects and other subjects is statistically significant; so are the differences in objects.

[2] $X^a = 12.84$, $p < .001$. Note that the difference in rate of particle ellipsis for animate and inanimate objects is not ;significant at the .01 level ($X^2 = 6.07$).

A Stochastic OT grammar which predicts these frequencies:

Definiteness

*WeakSu & Case 100.906
*StrongSu & Case 100.342
*StrongOj & Case 100.089
*Struc 99.380
*WeakOj & Case 99.282

	Rate of particle ellipsis per Fry 2001	Rate predicted by above grammar (rounded off to nearest percent)
N1 (Weak Subject)	.30	.30
N1 (Strong Subject) Pronoun or PN	.37	.37
N2 (Strong Object) Pronoun or PN	.41	.40
N2 (Weak Object)	.53	.51

*STRUC A STYLE SENSITIVE CONSTRAINT IN JAPANESE

CASUAL REGISTER

*WeakSu & Case	100.906
*StrongSu & Case	100.342
*StrongOj & Case	100.089
*Struc	99.380
*WeakOj & Case	99.282

WRITTEN REGISTER

{ *WeakSu & Case, *StrongSu & Case, *StrongOj & Case, *WeakOj & Case }

|

*Struc

(Boersma and Hayes 2001):

At the time of evaluation, the styleSensitivity value associated with *Struc will drive its selectionPoint down in the more formal register.

$selectionPoint_i = rankingValue_i + styleSensitivity_i \cdot Style + noise$

☞ Reduction in structure is associated with informal registers both in morphosyntax (Haiman 1985) and in phonology (Tranel 1999)

Conclusion to be drawn from Colloquial Japanese

It would be a mistake to conclude from the categorical nature of case marking in Written Japanese, that the grammar of Japanese lacks the constraint subhierarchies which characterize the likelihood of DSM and DOM.

When *Struc is *not* pushed down below the DCM constraints, the structure of the subhierarchies reveals itself in differential subject and object marking: more marked alignments are more likely to be case-lmarked than less marked ones.

GENERAL CONCLUSIONS

Differential Case Marking is shaped by two functional pressures:

- the need to distinguish subject and object
- the need to be economic

DCM shows clear evidence of 'markedness reversal':

- the properties which favor DOM are the inverse of those which favor DSM

Therefore, an account of DCM based on on $\mathcal{HA}$ of prominence scales and *Struc is appropriate.

- predicts the existence of both Differential Subject Marking for low prominence subjects and Differential Object Marking for high prominence objects ("markedness reversal").

- formally links DCM, a morphological phenomenon, to the avoidance of marked syntactic structure.

- provides a unified account of the cross-linguistic typology of categorical DCM and of language-particular variable DCM. The same implicational generalizations which structure the typological space across languages also structure frequency within individual languages.

- it can model the gradual expansion of DCM over time.

REFERENCES.

Boersma, P. and B. Hayes (2001). "Empirical tests of the Gradual Learning Algorithm." Linguistic Inquiry **32**(1): 45-86.

Fry, J. (2001). Ellipsis and *wa*-marking in Japanese conversation. Linguistics Department. Stanford, CA, Stanford University.

Haiman, J. (1985). Natural syntax: iconicity and erosion. Cambridge, Cambridge University Press.

Laca, B. (2000). "El complemento directo." Ms.

Lazard, G. (1982). "Le morphème râ en persan et les relations actancielles." Bulletin de la société de linguistique de Paris **73**(1): 177-208.

Müller, G. (1999). "Optionality in optimality-theoretic syntax." GLOT International **4:5**.

Ortmann, A. (to appear). Economy-based splits, constraints, and representations. More than words. A Festschrift for Dieter Wunderlich. I. Kaufmann and B. Stiebels. Berlin, Akademie-Verlag.

Stiebels, B. (2000). Linker inventories, linking splits, and lexical economy. Lexicon in focus. B. Stiebels and D. Wunderlich. Berlin, Akademie-Verlag**:** 211-245.

Stiebels, B. (2000). Typologie des Argumentlinkings: Ökonomie und Expresivität. Linguistics Department. Düsseldorf, Düsseldorf University.

Tranel, B. (1999). Optional schwa deletion: On syllable economy in French. Formal perspectives on Romance linguistics. J.-M. Authier, B. Bullock and L. Reed. Amsterdam, John Benjamins**:** 271-287.

Zeevat, H. and G. Jäger (2002). A reinterpretation of syntactic alignment. Proceedings of the 3rd and 4th International Symposium on Language, Logic and Computation. D. d. Jongh, H. Zeevat and M. Nilsenova. Amsterdam, ILLC.