DIFFERENTIAL CASE MARKING

OT Syntax and Typology

Summer School, Düsseldorf
Judith Aissen

Recall …

"…the most natural kind of transitive construction is one where the A is high in animacy and definiteness, and the P is lower in animacy and definiteness; and any deviation from this pattern leads to a more marked construction." (Comrie 1989, 128)

Q: What was the evidence for Comrie's claim …

A: The asymmetric distribution of more marked constructions.

In the domain of voice, we have already seen support for this implicational universal:

> If passive is obligatory when the Agent is *x*-Person and the Patient is *y*-Person, then passive is obligatory if the Agent is *z*-Person ($z < x$), or the Patient is *w*-Person ($w > x$).

There is also evidence in the domain of morphology. One rich source of support comes from Differential Case Marking.

What is Differential Case Marking?

Case marking systems in which some nominals with a given grammatical function $\mathcal{GF}$ are overtly case marked, but others are not.

- Differential object marking (DOM)

  A case marking system in which some objects, but not all, are overtly case marked.

- Differential subject marking (DSM)
  A case marking system in which some subjects, but not all, are overtly case marked.

DIFFERENTIAL OBJECT MARKING

Some examples:

Hebrew, in which *definite* objects are marked, but *not indefinite* ones (Givón 1978).

> Ha-seret her ʔa *ʔet-ha-milxama*.
> the-movie showed ACC-the-war
> The movie showed the war.
>
> Ha-seret her ʔa *(*ʔet-) milxama*.
> the movie showed (ACC-)war
> The movie showed a war.

Sinhalese, in which *animate* objects may be case marked, but *not inanimate* ones (Gair 1970).

> Mamə heṭə wæḍəṭə miniha evannan.
> I morrow work (dat) man send
> I will send the man to work tomorrow.
>
> Mamə heṭə wæḍəṭə miniha-wə evannan.
> I morrow work (dat) man-acc send
> I will send the man to work tomorrow.

Inanimate-referring nouns have no accusative form.

Romanian, in which object case-marking (*pe*) is

- Obligatory for animate-referring pronoun and proper noun objects.
- Optional (but preferred) for animate-referring, definite and indefinite specific objects
- Precluded for all inanimate-referring objects and for all non-specific objects.

(Farkas 1978; Dobrovie-Sorin 1994)

DOM is very common, but is realized in many different forms. Cross-linguistically, DOM varies…

- with respect to how case marking is realized;
- with respect to exactly which objects can be case marked;
- and with respect to whether marking is obligatory or optional.

Is DOM a unified phenomenon? Yes…

THE GENERALIZATION UNDERLYING DOM

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

[Silverstein, 1976 #254; Comrie, 1979 #62; Comrie, 1980 #64; Comrie, 1986 #65; Comrie, 1989 #66; Lazard, 1982 #621; Lazard, 1984 #620; Bossong, 1985 #651; Bossong, 1991 #650]

Prominence is assessed on two scales:

Animacy Scale:

Human > Animate > Inanimate

Definiteness Scale:

Pronoun > PN > Definite > Indefinite Specific > NonSpecific

Hebrew, in which *definite* objects are case marked, but *not indefinite* ones.

Sinhalese, in which *animate* objects may be case marked, but *not inanimate* ones.

Romanian, in which object case marking is

- Obligatory for animate-referring pronoun and proper noun objects.
- Optional (but preferred) for animate-referring, specific objects.
- Precluded for inanimate-referring and non-specific objects.

Languages vary w which of the two scales is relevant for DOM, t wrt the 'cut-off' point for DOM, but they all appear to be consistent with the "Generalization Underlying DOM".

WHAT UNDERLIES THE GENERALIZATION?

- Why are the animacy scale and the definiteness scale organized as they are?

- Why would object animacy and definiteness favor differential case marking?

  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.

"…the most natural kind of transitive [frequent, JA] construction is one where the A is high in animacy and definiteness, and the P is lower in animacy and definiteness; and any deviation from this pattern leads to a more marked construction." (Comrie 1989, 128)

Some evidence from frequency:

Swedish (Zeevat and Jäger 2002)

p(subj|NP) p(ojb|NP)	77% 23%
p(subj|hum)	97%
p(subj|ego)	97%
p(obj|-def)	87%
p(obj|inan)	46%

→ DOM counteracts the BIAS which favors parsing animate, definite nominals as SUBJECTS (Zeevat and Jäger 2002).

An account of DOM in terms of Harmonic Alignment (*HA*)

The critical constraints involve alignment of *GF* with animacy and definiteness.

HA (GF, Animacy Scale)	
Harmonic Alignments	Su/Hum ≻ Su/Anim ≻ Su/Inan Oj/Inan ≻ Oj/Anim ≻ Oj/Hum
Constraint Subhierarchies	*Su/Inan » *Su/Anim » *Su/Hu *Oj/Hum » *Oj/Anim » *Oj/Inan
HA (GF, Definiteness Scale)	
Harmonic Alignments	Su/Pro ≻ Su/PN ≻ Su/Def ≻ Su/Indef ≻ Su/Nspec Oj/NSpec ≻ Oj/Indef ≻ Oj/Def ≻ Oj/PN ≻ Oj/Pro
Constraint Subhierarchies	*Su/NSpec »*Su/Indef »*Su/Def »*Su/PN »*Su/Pro *Oj/Pro»*Oj/PN » *Oj/Def » *Oj/Indef » Oj/NSpec

These constraints are motivated independent of DOM, e.g.

> Chamorro, where human patients cannot be realized as objects if the agent is inanimate (Chung 1981; Cooreman 1987; Chung 1998)
>
> *Su/Inan & *Oj/Anim (penalizes active, forces passive)
>
> Tagalog, where definite patients cannot be realized as objects (Foley and Van Valin 1984; Kroeger 1993)
>
> *Oj/Pro»*Oj/PN » *Oj/Def (force 'passive')

☞ The same constraint subhierarchies should be used to describe *both avoidance and marking* of high prominence objects.

DEVELOPING AN ACCOUNT OF DOM BASED ON $\mathcal{HA}$

1. The constraints from the above table penalize objects of various types. But here it is not prominent objects which are avoided, but prominent objects which are not case-marked.

2. The relevant constraint is:

   Case: violated if a nominal has no value for the feature CASE

   (Case =*ø_{case} of (Aissen 1999; Aissen 2000)

3. To express the fact more prominent objects are most likely to be case-marked, Case is locally conjoined with Oj-indexed subhierarchies.

The ranking of the source subhierarchy is maintained:

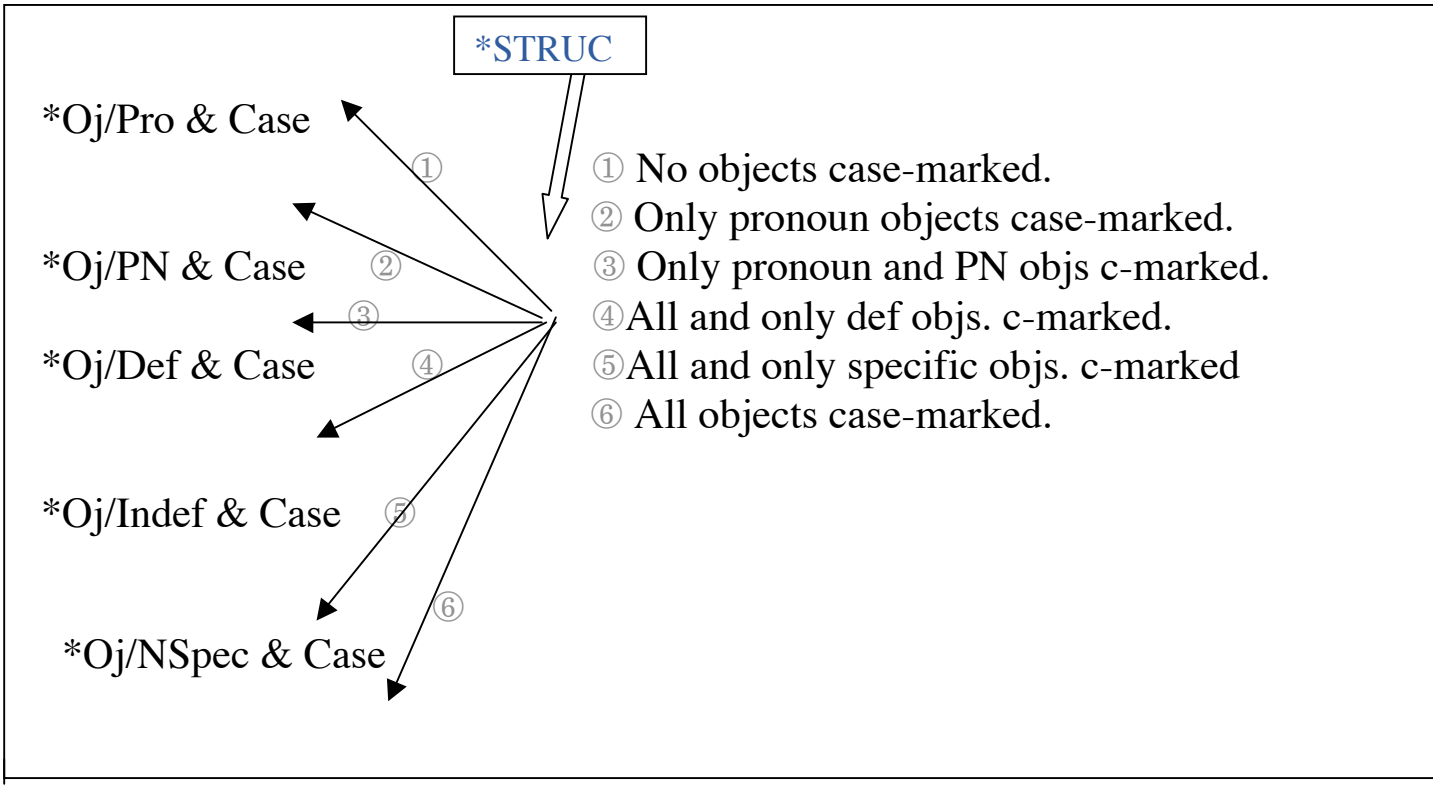
Local conjunction of Case with the subhierarchy on object animacy	**Local conjunction of Case with the subhierarchy on object definiteness**
*Oj/Hum & Case »	*Oj/Pro & Case »
*Oj/Anim & Case »	*Oj/PN & Case »
*Oj/Inan & Case	*Oj/Def & Case »
	*Oj/Indef & Case »
	*Oj/NSpec & Case

These constraints penalize absence of case marking. What penalizes its presence? Economy…

*STRUC: penalizes a value for the morphological category CASE

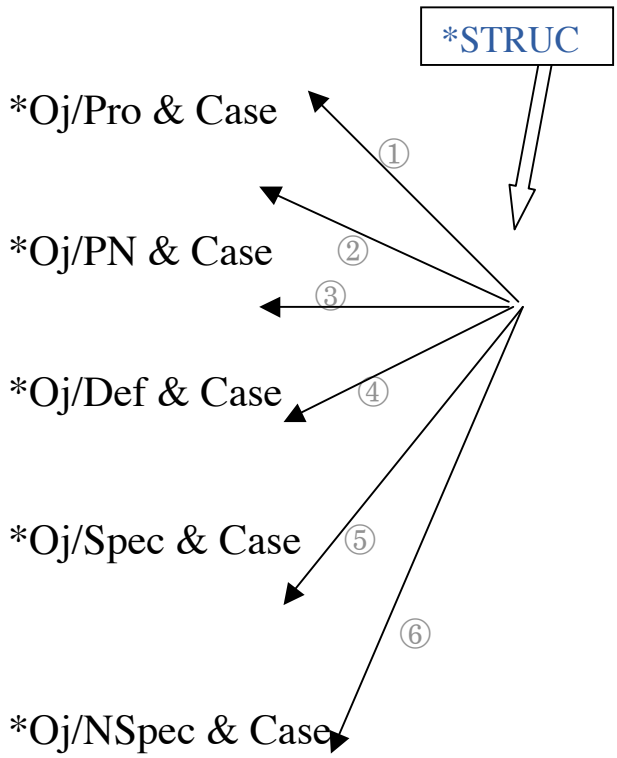
The ranking of *STRUC with respect to the 'iconicity' subhierarchies determines how much overt case marking there is in DOM systems. The higher it is ranked, the fewer object-types will be case marked.

DOM BASED ON DEFINITENESS



Typological prediction for categorical DOM systems based on definiteness: In a language L, if objects at some rank on the definiteness hierarchy are obligatorily case marked, then objects at all higher ranks will also be obligatorily case marked.

Is this prediction correct? It appears to be.



① Kalkatungu: no objects case-marked [Blake, 1979 #725].

② Catalan: only pronouns case-marked

③ Pitjantjatjara: only pronouns and PNs case-marked [Bowe, 1990 #663].

④ Hebrew: all and only definites case-marked [Givón, 1978 #660].

⑤ Turkish: all and only indefinite specifics case-marked [Enç, 1991 #532].

⑥ Written Japanese, Dhalandji: all objects case-marked [Austin, 1981

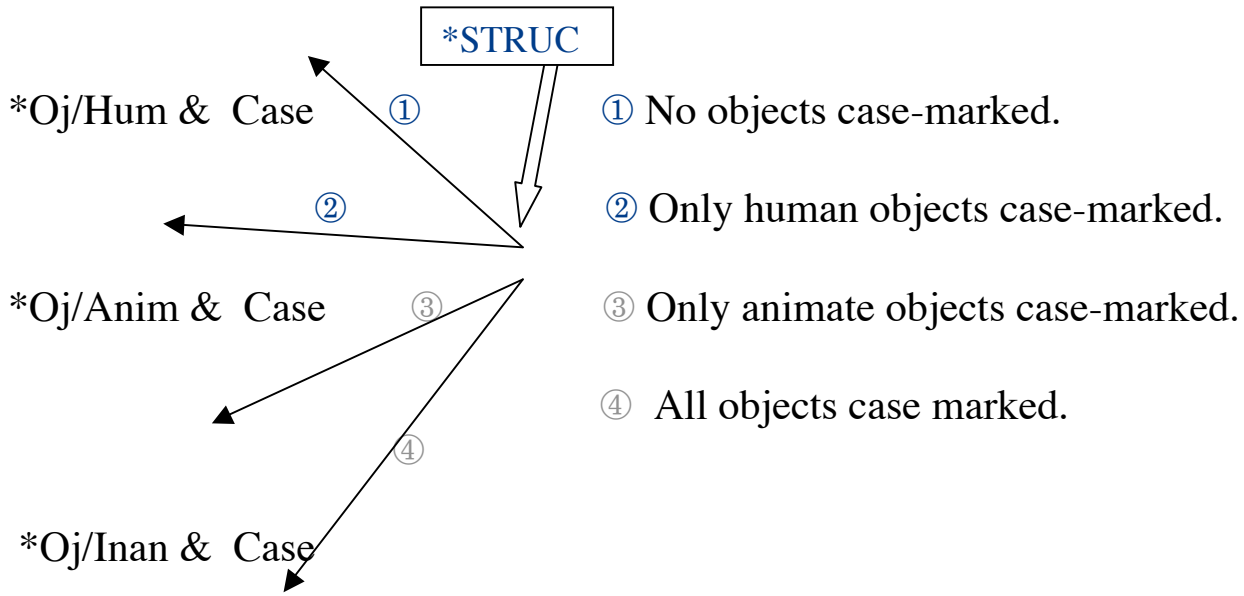
Hebrew

Patient: specific indefinite	*Oj/Def & Case	*STRUC$_C$	*Oj/Spec & Case	*Oj/NSpec & Case
Oj / specific indefinite CASE: ACC		*!		
☞Oj / specific, indefinite CASE:			*	

Turkish

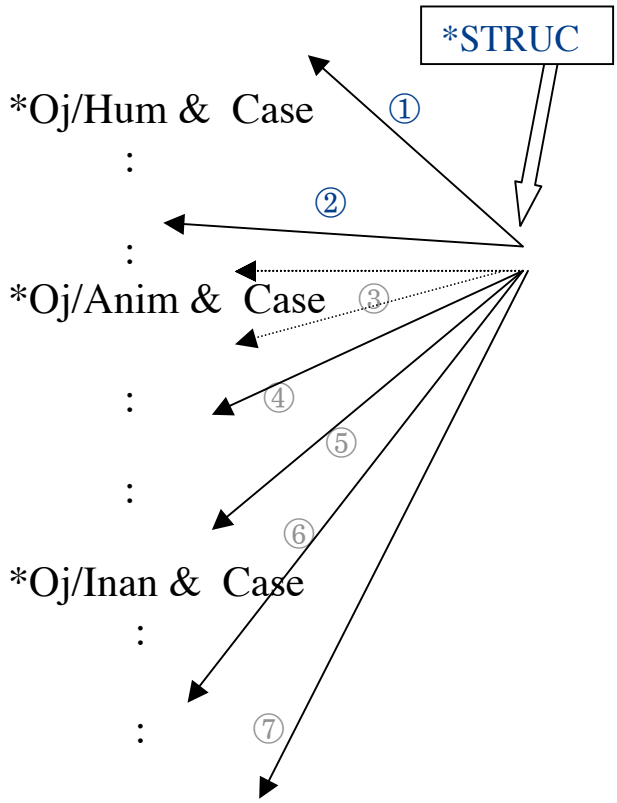
Patient: specific indefinite	*Oj/Def & Case	*Oj/Spec & Case	*STRUC$_C$	*Oj/NSpec & Case
☞Oj/ specific indefinite CASE: ACC			*	
Oj /specific, indefinite CASE:		*!		

DOM BASED ON ANIMACY



Typological prediction for categorical DOM systems based on animacy: In a language L, if objects at some rank on the animacy hierarchy are obligatorily case marked, then objects at all higher ranks will also be obligatorily case marked.

Are these predictions correct? They seem to be, though we need to recognize that the animacy scale is more articulated than indicated so far, that it is subject to some language-particular ranking, and that the cut-off point for DOM need not fall at the 'big' boundaries.



① Kalkatungu: no objects case-marked.

② Yiddish: only some human objects case-marked.

③ Sinhalese: all animates optionally case-marked.

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

⑤ Dhargari: all animate objects case marked.

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

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

INTERIM SUMMARY

The analysis expresses the basic generalization underlying DOM:

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

How does it do so?

The basic tension is between

Economy -- *Struc, which penalizes morphological case, and
Iconicity -- the subhierarchies *Oj/Hum & Case » …
*Oj/Pro & Case » …

which penalize the absence of morphological case in marked configurations more strenuously than in unmarked configurations.

VARIABLE DOM (MORE BELOW)

So far, we have focused on categorical cases. But DOM exists in variable form as well, e.g. Sinhalese.

To characterize variable DOM, we assume the Stochastic Generalization of Optimality Theory.

E.g.

Idealized Sinhalese

Outputs:	Case	NoCase
Anim	50%	50%
Inan	0%	100%

A constraint ranking:

*Oj/Anim & Case	105
*Struc	105
*Oj/Inan & Case	90

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

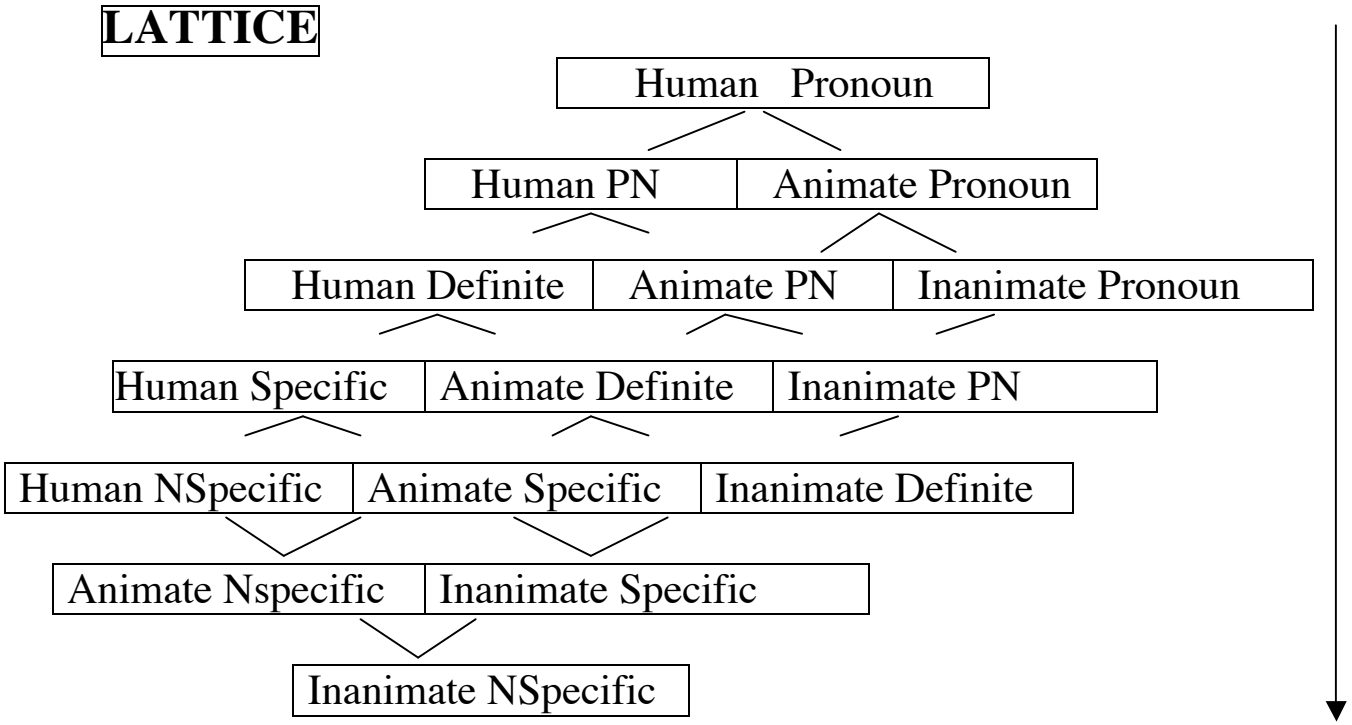
Recall Romanian, where object case-marking is

- Obligatory for animate-referring pronoun and proper noun objects.
- Optional (but preferred) for animate-referring, definite objects
- Precluded for inanimate-referring and non-specific objects.

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

How is prominence computed when two dimensions are involved?

Take the cross-product of the two scales.



Intuitively, DOM flows from the top of the lattice down ↓.

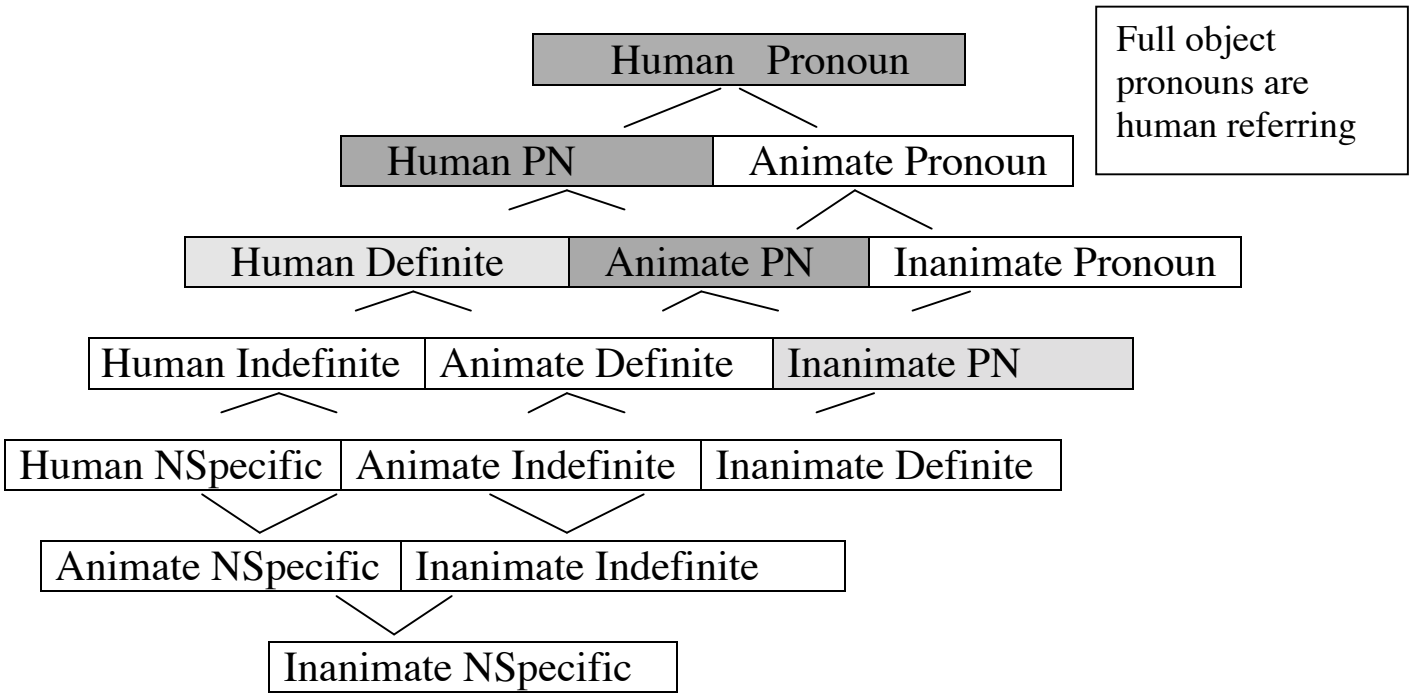
Def: □ is more prominent than □ iff □ dominates □.

Two-dimensional DOM Universals:

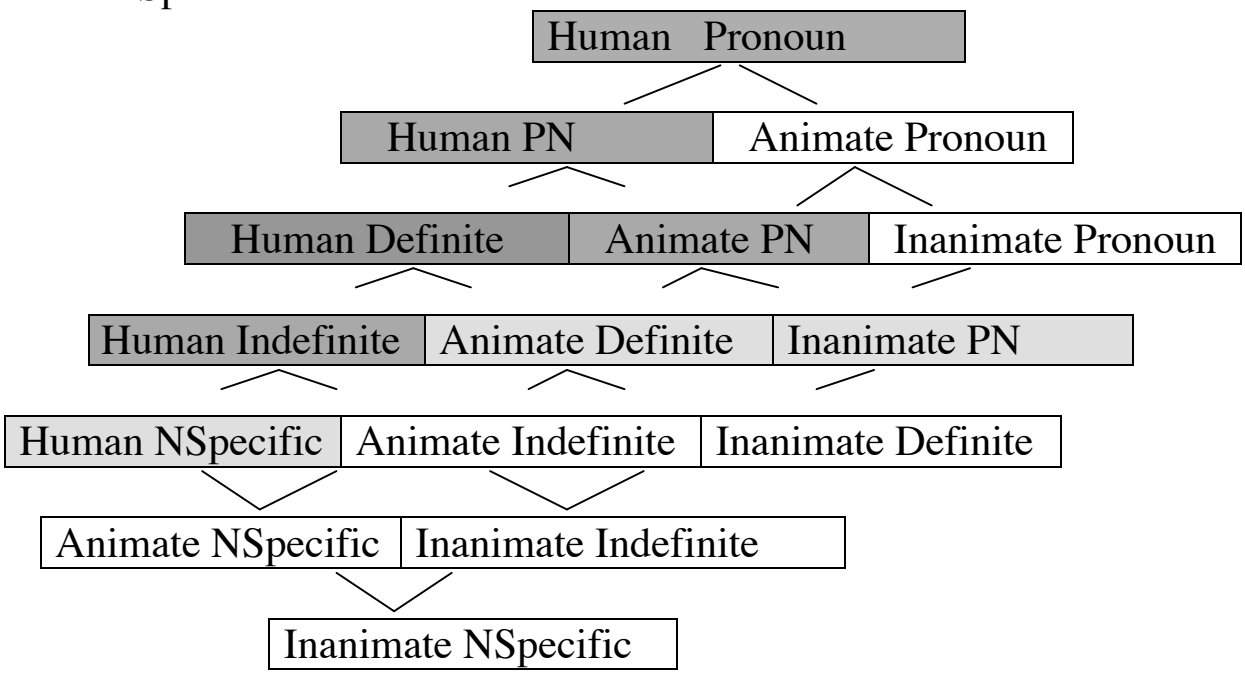
1 If □, an object, can be overtly case-marked, then (all else equal) an object more prominent than □ can be overtly case-marked.
2 If □, an object, must be overtly case-marked, then (all else equal) an object more prominent than □ must be overtly case-marked.

=> Partial ranking: Hebrew vs. Romanian

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

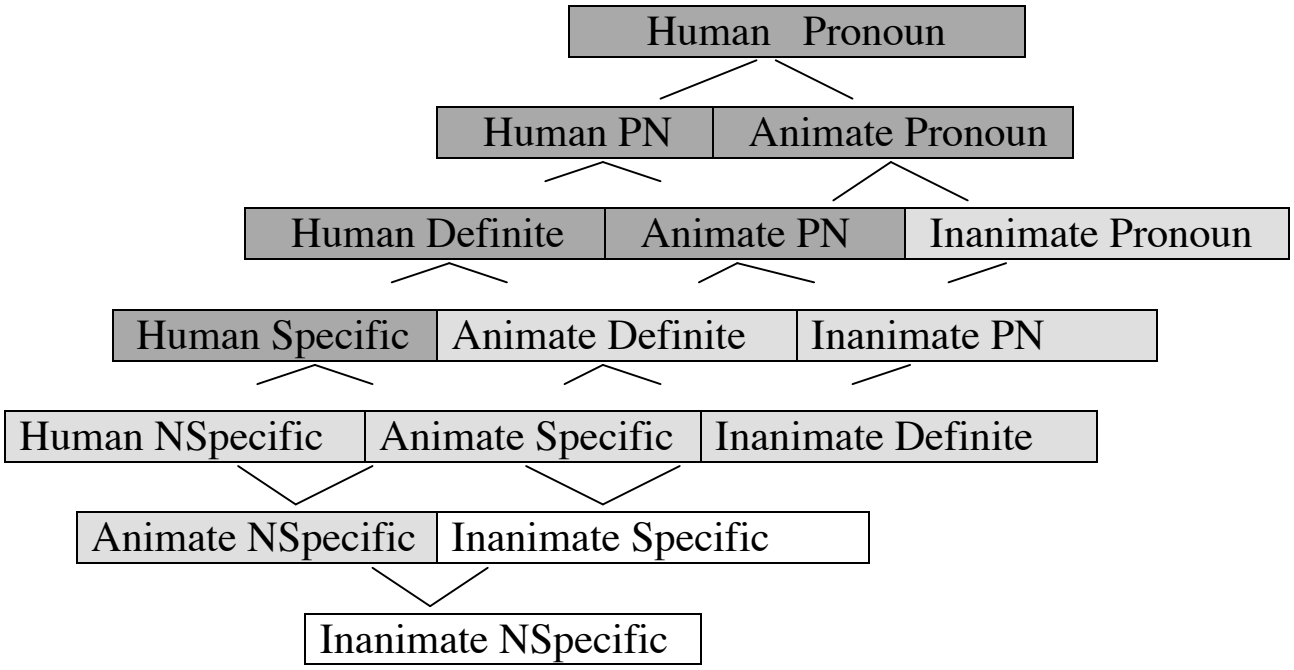


Modern Spanish

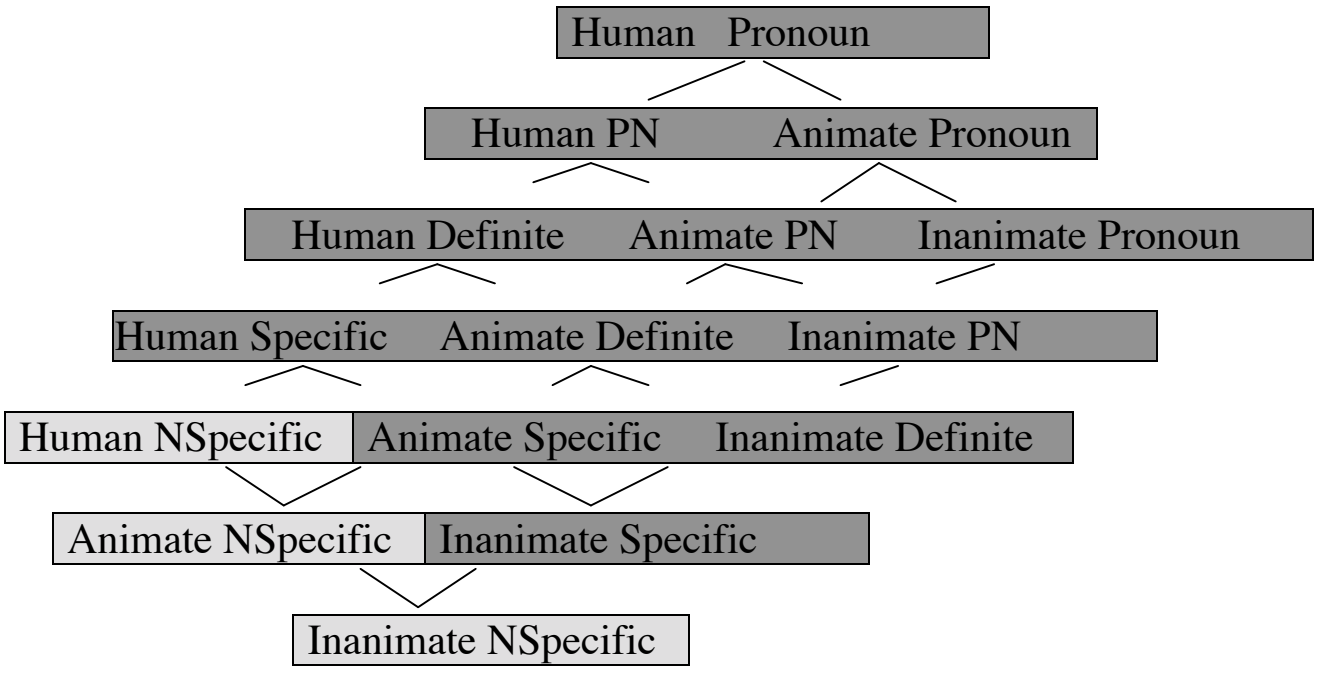


[1] Dark cells: obligatory case marking; light cells: optional case marking; white cells: no case marking.

DOM in Hindi

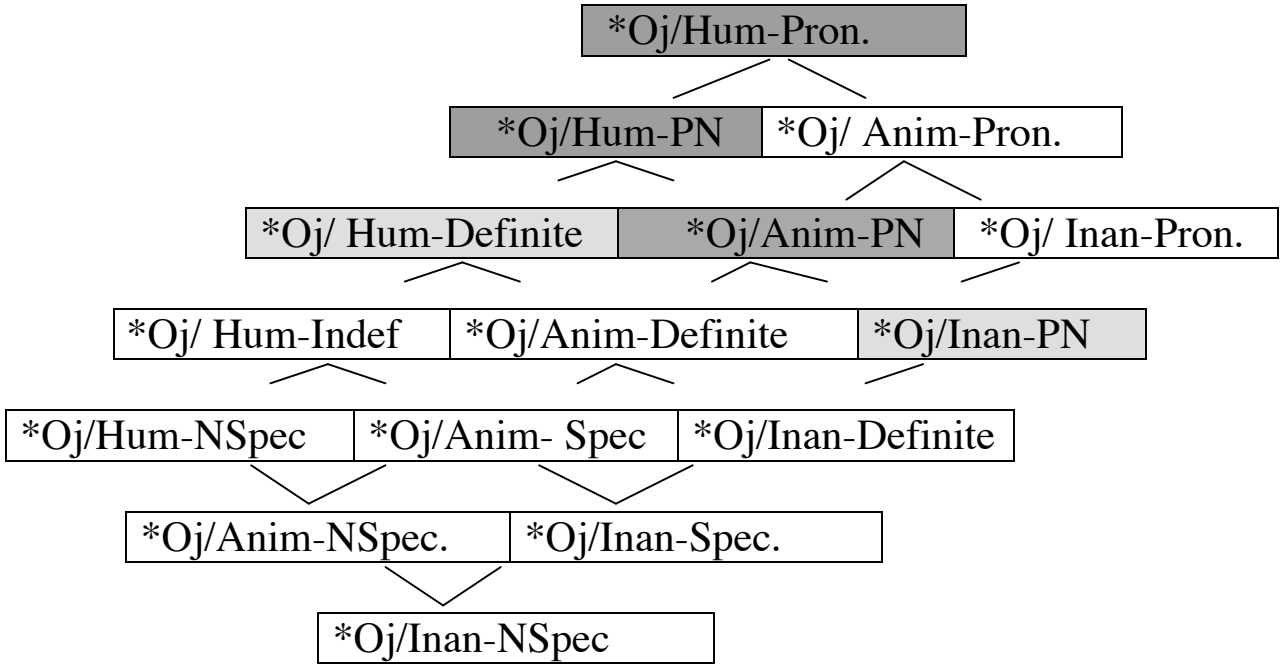


DOM in Persian



Characterizing two-dimensional DOM systems

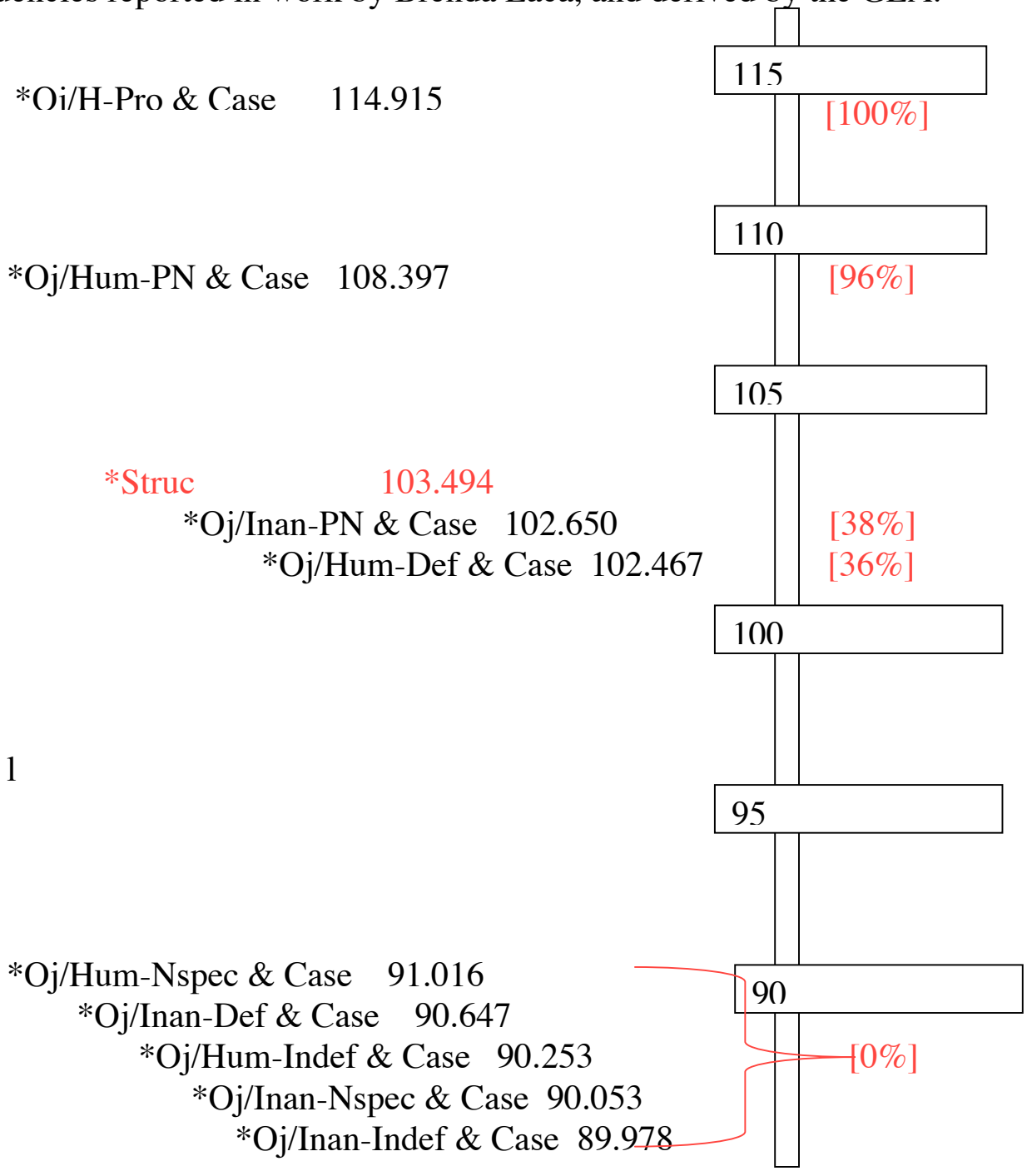
A partially ordered constraint set isomorphic to the above lattice is derived by local conjunction of the constraints on object definiteness and object animacy and further conjunction with Case (Case is omitted below).
[Read: **Oj/Hum-Pron.* as [[*Oj/Human $\&_{DP}$ *Oj/Pronoun] & Case]



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.

A Stochastic OT grammar of 12th century Spanish DOM, based on frequencies reported in work by Brenda Laca, and derived by the GLA.



RETURNING TO SOME EARLIER QUESTIONS ABOUT $\mathcal{HA}$

Q: $\mathcal{HA}$, by its nature, produces constraints which penalize both marked and unmarked structure (high-ranked and low-ranked constraints, resp.) Are the latter (e.g. *Oj/Indefinite) necessary? Could they be dropped (as suggested, for example, in Zeevat and Jäger 2002)?

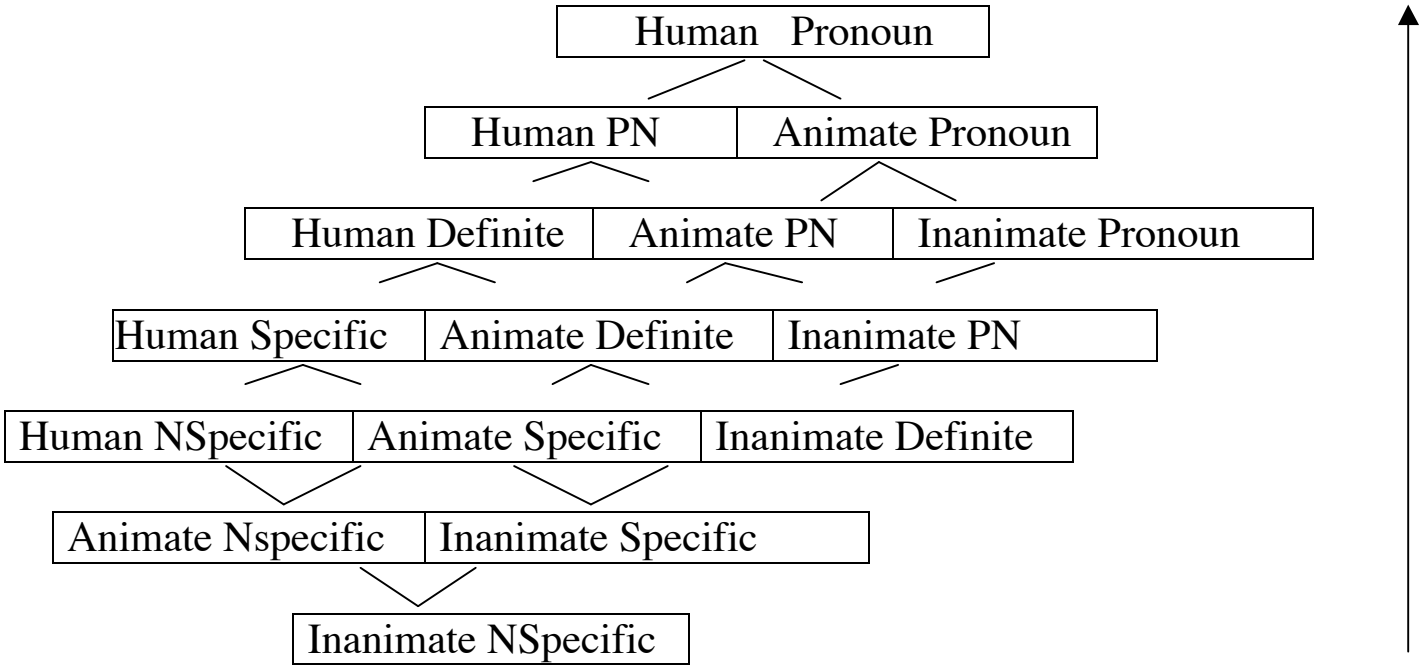
A: Not if this account of DOM is correct. Indefinite objects are unmarked (relative to definite ones). Yet there are languages in which they must be case marked (e.g. Turkish). Inanimate objects are relatively unmarked (relative to animate ones). Yet they may be case-marked in Hindi, for example, albeit only when definite. Characterizing these systems as above requires constraints that penalize relatively unmarked structure.

Q: $\mathcal{HA}$, by its nature, simultaneously characterizes relative markedness of associations with both ends of the binary scale. That is, for each Subject-oriented subhierarchy , $\mathcal{HA}$ automatically produces Object- and Oblique-oriented subhierarchies. Do we need both of the constraint subhierarchies that $\mathcal{HA}$ generates?

A. Yes, if this account of DCM is on the right track.

Object-oriented constraints are needed to characterize DOM. Subject oriented ones would be irrelevant. The existence of subject-oriented constraints predicts that DSM should exist. It does.

Differential Subject Marking



3 If □, a subject, can be overtly case-marked, then (all else equal) a subject *less* prominent than □ can be overtly case-marked.

4 If □, a subject, must be overtly case-marked, then (all else equal) a subject less prominent than □ must be overtly case-marked.

(Intuitively, Differential Subject Marking percolates from the bottom of the lattice to the top ↑.) Examples:

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

Guugu Yimidhirr, in which all lexical NPs are overtly case marked, but personal pronouns are not.

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

Yalarnga, Rembarnga, in which all subjects are marked (and no objects).

OVERALL TYPOLOGY OF SUBJECT AND OBJECT CASE MARKING

Pure accusative system:

*Oj/$\mathcal{X}$ & Case » *Struc » *Su/$\mathcal{X}$ & Case

Pure ergative system:

*Su/$\mathcal{X}$ & Case » *Struc » *Oj/$\mathcal{X}$ & Case

Differential object marking:

*Struc interpolated among the *Oj/$\mathcal{X}$ & Case constraints

Differential subject marking:

*Struc interpolated among the *Su/$\mathcal{X}$ & Case constraints

Q: This may be a reasonable way to characterize the typology of DCM. But what evidence is there that the full hierarchy of constraints is present in the grammars of individual languages (or individual speakers)? Why appeal to a full hierarchy when what needs to be drawn in individual languages is simply a binary (or trinary) distinction?

A: The same structure which underlies the cross-linguistic typology of DCM systems underlies variable DCM. Languages with variable DCM 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.

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)

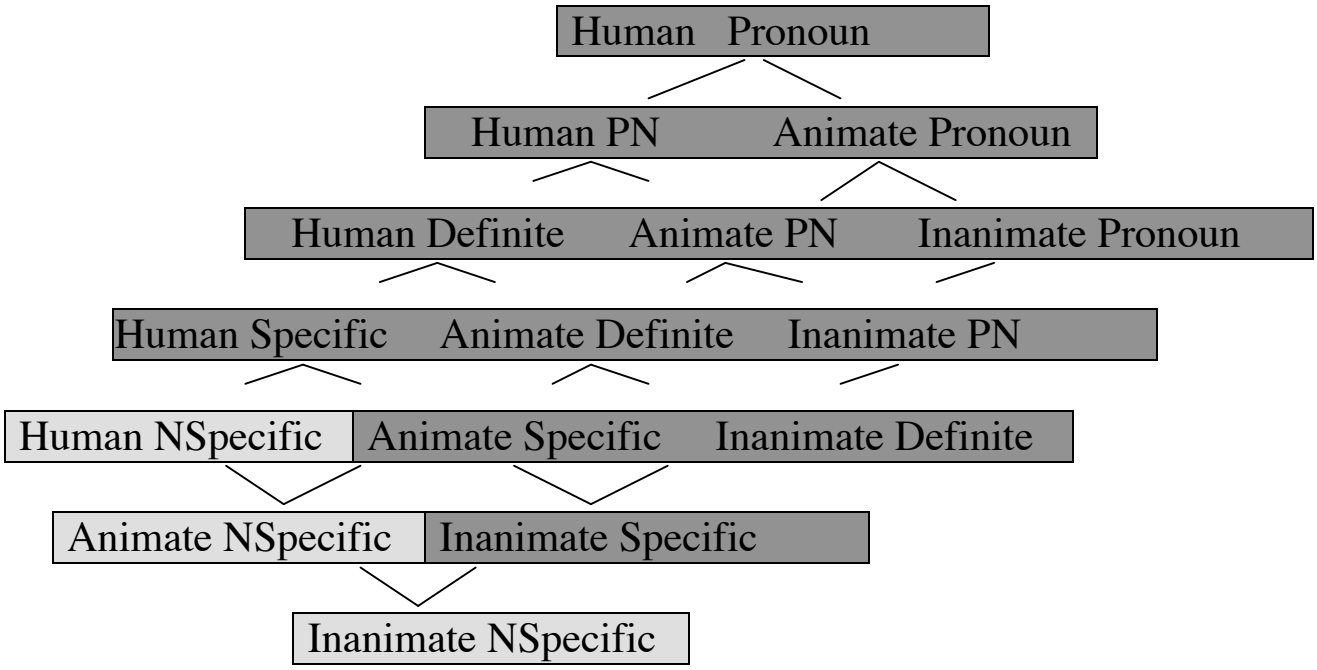
Indefinites are optionally marked, but according to Lazard, specific indefinites are, as a rule, marked. He notes two classes of specific indefinites: those which have a partitive sense, and those with the sense of *a certain*. Both classes require the suffix *–râ*.

> 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

VARIABLE DIFFERENTIAL CASE MARKING – COLLOQUIAL JAPANESE

Subject and object postpositions (*ga*, *o*) are required in Written Japanese. But in informal spoken Japanese, they may be dropped.

The following data is from (Fry 2001), which was based on the annotated CallHome Japanese corpus

N1 = subject, N2 = direct object

Following particle?	Animate				Not animate			
	N1		N2		N1		N2	
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)

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

Following particle?	Proper Noun or personal pronoun				Other			
	N1		N2		N1		N2	
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.

GENERAL CONCLUSIONS

A Stochastic OT account of Differential Case Marking based on $\mathcal{HA}$ of prominence scales:

- predicts the cross-linguistic typology of DCM, and characterizes language particular DCM systems. Also predicts full accusative systems (all objects marked, no subjects) and full ergative systems (all transitive subjects marked, no objects).

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

- predicts the existence of both Differential Subject Marking for low prominence subjects and Differential Object Marking for high prominence objects.

- 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 cross-linguistically structure frequency within individual languages.

REFERENCES.

Aissen, J. (1999). "Markedness and subject choice in optimality theory." Natural Language and Linguistic Theory **17**(4): 673-711.

Aissen, J. (2000). Differential object marking: Iconicity vs. economy.

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

Chung, S. (1981). Transitivity and surface filters in Chamorro. Studies in Pacific languages and cultures, in honour of Bruce Biggs. J. Hollyman and A. Pawley. Aukland, Linguistic Society of New Zealand**:** 311-332.

Chung, S. (1998). The design of agreement. Evidence from Chamorro. Chicago, University of Chicago Press.

Comrie, B. (1989). Language universals and linguistic typology. Chicago, University of Chicago Press.

Cooreman, A. (1987). Transitivity and discourse continuity in Chamorro narratives. Berlin, Mouton de Gruyter.

Dobrovie-Sorin, C. (1994). The syntax of Romanian. Berlin, Mouton de Gruyter.

Farkas, D. (1978). Direct and indirect object reduplication in Romanian. Papers from the fourteenth regional meeting of the Chicago Linguistic Society. Chicago, Chicago Linguistic Society**:** 88-97.

Farkas, D. (1997). Towards a semantic typology of noun phrases. Colloque de syntaxe et sémantique de Paris, Université Paris 7.

Foley, W. and R. Van Valin (1984). Functional syntax and universal grammar. Cambridge, Cambridge University Press.

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

Gair, J. (1970). Colloquial Sinhalese clause structures. The Hague, Mouton.

Givón, T. (1978). Definiteness and referentiality. Universals of human language. J. Greenberg. Stanford, CA, Stanford University Press. **4:** 291-330.

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

Kroeger, P. (1993). Phrase structure and grammatical relations in Tagalog. Stanford, CA, CSLI Publications.

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

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.