Rational communication explains Differential Argument Marking and typological asymmetries between subject and object marking

Jesse Holmes (jesse.daniel.holmes@ut.ee)[1]
[1]Institute of Estonian and General Linguistics, University of Tartu, Estonia

Abstract

Transitive events involve an agent (source of the action) and a patient (recipient of the action), and the successful communication of such events requires conveying these roles. Languages employ various strategies for this, including word order and case marking. Differential Argument Marking (DAM), denoting systems in which case marking only occurs in certain contexts, has garnered considerable attention for its cross-linguistic prevalence. One account posits that DAM arises via efficiency: referents with properties atypical for their role (and therefore less predictable) create pressure for marking, while typical referents do not. This predicts similar numbers of languages with atypical agent-marking and atypical patient-marking. However, patient-marking is not only more common, but agent-marking systems often violate expected typicality patterns. To address this asymmetry, we develop a computational model showing that it emerges naturally from the interaction of typicality biases with production costs.

Keywords: differential argument marking; communicative efficiency; computational modeling; linguistic typology; case marking; word order

Introduction

Transitive events involve two participants: an agent, who initiates the action, and a patient, who is affected by it. Successful communication of such events requires conveying which participant fills which role. Languages employ various strategies to achieve this, including fixed word order and morphological case marking. Of particular interest are systems of Differential Argument Marking (DAM), in which case marking occurs only in certain contexts rather than uniformly across all instances of a grammatical relation (Aissen, 2003; Bossong, 1991). The most widely studied instantiation is Differential Object Marking (DOM: sentences 1a-b), in which a subset of objects receive overt case marking based on properties such as animacy or definiteness. Differential Subject Marking (DSM: sentences 2a-b), including split ergativity and optional ergative case, is the analogous phenomenon for subjects (de Hoop & de Swart, 2007).

(1) Spanish: DOM conditioned by animacy (Haspelmath, 2021)

a. Veo a la mujer.
see.PRS.1SG OBJ the woman
'I see the woman.'

b. Veo la casa.
see.PRS.1SG the house
'I see the house.'

(2) Fore: DSM conditioned by animacy (Scott, 1978)

a. Yagaa-wama wá aegúye
pig-SUBJ man hit
'The pig hits the man.'

b. Yagaa wá aegúye.
pig man hit
'The man hits (or kills) the pig.'

The prevalence and systematicity of DAM patterns across unrelated language families has motivated substantial theoretical interest. One influential account holds that DAM reflects communicative efficiency: referents with properties atypical for their role are less predictable and therefore create pressure for overt marking, while typical referents can remain unmarked because the listener can infer their role from context (Aissen, 2003; Haspelmath, 2021; Levshina, 2021). For example, in sentences (1a-b), the human object *mujer* is marked with *a* because animate patients are atypical, while inanimate *casa* is typical and unmarked; in (2a-b), the agent is marked when less animate (and less typical). This account generates a clear prediction: if atypicality drives marking, we should observe similar cross-linguistic frequencies of atypical agent-marking and atypical patient-marking. However, Malchukov (2008) documented that patient-marking systems are not only more common than agent-marking systems, but agent-marking often fails to follow expected typicality patterns (see below). This asymmetry poses a puzzle for efficiency-based accounts.

In this paper, we develop a computational model that addresses this puzzle. Following the methodology of rational analysis (Anderson, 1990; Marr, 1982), we specify the abstract communicative problem facing speakers and listeners and derive what an optimal solution must look like. We show that the DOM/DSM asymmetry emerges naturally from the interaction of typicality-based interpretation biases with production costs, without requiring additional stipulation.

Theories on the emergence of DAM

Several theoretical traditions have addressed the functional motivations for DAM. The oldest and most influential is the distinguishability hypothesis, which holds that case marking serves to identify grammatical relations when they might otherwise be ambiguous (Comrie, 1989; Dixon, 1994). On this view, overt marking is deployed where confusion between agent and patient is most likely.

Aissen (2003) formalized a related intuition within Optimality Theory as a competition between iconicity, which favors marking atypical arguments, and economy, which penalizes overt morphology. Objects that are animate or definite deviate from the prototype for patients and therefore attract marking; objects that are inanimate and indefinite conform to expectations and can remain unmarked.

Recent work has reframed DAM in terms of communicative efficiency more broadly. Levshina (2021) argued that differential marking reflects optimization for efficient communication, with marking deployed where it provides the greatest informational benefit relative to its cost. Smith and Culbertson (2025) provided experimental evidence that typicality-based DAM patterns emerge from communicative interaction rather than learning biases: participants using an artificial language with optional case marking produced differential patterns only when communicating with a partner, not in non-communicative recall tasks. This suggests that languages adapt to communicative efficiency constraints as a result of the communicative problem itself, rather than through human learning biases favoring efficient structures.

A persistent puzzle for these accounts (and previous modeling work; see Jäger, 2007) is the asymmetry between DOM and DSM. Malchukov (2008) documented that DOM is not only more common than DSM but also more consistent in its conditioning. DOM systematically marks animate, definite, or given objects, in line with the prediction that atypical patients receive overt marking. DSM, by contrast, shows inconsistent patterns: some languages mark atypical subjects (inanimate or indefinite agents), while others mark typical subjects (animate, pronominal agents), and many show complex splits that do not reduce to a single functional principle. Malchukov attributed this asymmetry to conflicting functional pressures on subject marking, but the source of this conflict has not been formally characterized.

Cognitive and communicative foundations

The model we develop treats the assumptions underlying DAM not as arbitrary stipulations but as rational responses to the computational problems faced by speakers and listeners. The central claim is that the interaction of these individually rational strategies gives rise to differential marking patterns, including the asymmetry between DOM and DSM.

The listener's problem

A listener receiving a signal about a transitive event faces uncertainty about which referent is the agent and which is the patient. The computational problem is to recover the intended meaning as quickly and accurately as possible given limited processing resources. Two strategies are rational responses to this problem.

First, relying on prior expectations is efficient when signals are ambiguous or incomplete. Bayesian accounts of cognition formalize this intuition: an optimal agent combines prior beliefs with incoming evidence to form posterior estimates (Chater & Oaksford, 2008). In the context of thematic role assignment, this means expecting typical role configurations more than atypical ones, and interpreting ambiguous input accordingly. Priors reduce the effort required to interpret expected messages and allow reasonable inferences even before all information has arrived.

Second, incremental processing is efficient because it allows interpretation to begin before the signal is complete, reducing memory load and enabling faster response. If the listener updates beliefs after each word rather than waiting for the full utterance, ambiguity can potentially be resolved earlier.

Both strategies are empirically well supported. Comprehenders generate predictions about upcoming material based on context, with violations producing measurable neural and behavioral responses (Federmeier, 2007; Kuperberg & Jaeger, 2016). Research on thematic role processing shows that listeners rapidly generate expectations about argument properties based on verbs and prior context (Bickel, Witzlack-Makarevich, Choudhary, et al., 2015; McRae et al., 1998) and rely on typicality biases in interpreting ambiguous sentences (Holmes & Vihman, 2025). Evidence for incremental processing is similarly robust: eye-tracking studies demonstrate that listeners integrate information in real time, making anticipatory fixations within hundreds of milliseconds (Altmann & Kamide, 1999; Tanenhaus et al., 1995).

The speaker's problem

A speaker who has observed a transitive event must produce an utterance enabling the listener to recover the correct role assignment. The computational problem is to communicate successfully while minimizing production effort. Two considerations follow.

First, minimizing production cost is rational because articulation requires time and energy. If two utterances are equally informative, the shorter or easier one should be preferred.

Second, the speaker must balance this pressure against the listener's needs. A rational speaker considers what the listener expects and deviates from low-cost defaults when doing so substantially improves interpretability. This is the core insight of Gricean pragmatics (Grice, 1975) and its formal descendants (Frank & Goodman, 2012): speakers design utterances for addressees who are themselves reasoning about speaker intentions.

These assumptions are also grounded empirically. Bock and Warren (1985) demonstrated that conceptual accessibility influences word order, with more accessible referents appearing earlier, and Tal et al. (2022) revealing similar effects of givenness on word order in non-communicative production tasks. MacDonald (2013) synthesized such literature, arguing that production biases systematically shape the distribution of linguistic forms. Evidence for informativity-cost tradeoffs comes from studies showing that speakers produce optional material more often when it aids disambiguation (Jaeger, 2010).

Agent salience

A key feature of transitive events is that agents and patients differ in cognitive salience. Agents are the causal initiators of events: knowing what the agent did often allows one to predict the patient's state, while the reverse is less true. Attending preferentially to agents is therefore a rational strategy for learning about causal structure. Relatedly, agents—particularly animate agents capable of self-initiated action—are potential sources of threat or opportunity, and attention to them may reflect selective pressures favoring detection of potential actors in the environment (New et al., 2007).

We take agent salience to be cognitively prior to communication rather than derived from it: it reflects evolved attentional priorities and the predictive structure of events, which the communicative system then exploits. The privileged status of agents is well documented (e.g., Meir et al. 2017): speakers' eye movements during event description tend to fixate agents before patients (Gleitman et al., 2007); Isasi-Isasmendi et al. (2023) review evidence for an agent preference in visual event apprehension; and Cohn and Paczynski (2013) finds that agents in visual narratives elicit greater predictive processing and longer viewing times than patients. Our model therefore takes the salience–role correlation as an exogenous input, and asks what marking patterns follow when efficient communication operates over it.

From individual strategies to typological patterns

The strategies outlined above characterize rational responses to the computational problems of individual speakers and listeners. Our model shows that when these strategies interact in a communicative system, differential marking emerges. The DOM/DSM asymmetry follows from how typicality biases interact with production costs asymmetrically for objects versus subjects. Because agents are typically salient and patients typically are not, the pressure to mark atypical configurations differs systematically between the two argument types. We develop this analysis formally in the following section.

Rational model of communication

We now develop a computational model that formalizes the cognitive and communicative foundations outlined above. Following the methodology of rational analysis (Anderson, 1990; Marr, 1982), we specify the communicative problem abstractly, characterize the relevant environmental and cognitive structure, and derive what an optimal solution must look like. The model analyzes a minimal communication system to isolate which patterns are demanded by the problem structure.

Meanings, contexts, and utterances

Consider a speaker who observes a transitive event involving two referents x and y, one acting as agent A and one as patient P. The speaker must produce an utterance enabling a listener to recover the correct role assignment. We formalize a meaning as a mapping m from referents r to roles θ, $m = \{r \mapsto \theta \mid r \in \{x, y\}, \theta \in \{A, P\}\}$, specifying which referent fills which role.

The referents do not exist in isolation but are situated within the speaker's cognitive state, where they may be differentially activated. We define salience s_r as a referent's scalar activation level, encompassing information-structural notions such as givenness, topicality, and attentional focus.[1] For any pair of referents, one will typically be more salient than the other, yielding a context space $c \in \{s_x > s_y, s_y > s_x\}$.

As discussed in the previous section, salience correlates with thematic role: agents, as causal initiators, tend to command more attention than patients. We formalize this as a likelihood ratio β_s representing how much more likely the higher-salience referent is to be the agent:

$$P(r \mapsto A, r' \mapsto P \mid s_r > s_{r'}) = \frac{\beta_s}{\beta_s + 1} \tag{1}$$

The magnitude of β_s reflects the activation differential between typical agents and patients. We set this parameter to $\beta_s = 2$, corresponding to a one-bit difference in activation. This means the higher-salience referent is twice as likely to be the agent as the patient. This choice is the simplest non-trivial assumption that generates asymmetric predictions.

Speakers encode meanings using two devices: word order $\omega \in \{(x, y), (y, x)\}$, specifying which referent is produced first, and morphological marking $\mu \in \{\sigma_A, \sigma_P, \emptyset\}$,[2] optionally indicating the role of a referent. Re-expressing everything in role-relative terms (substituting roles for referents given the intended meaning), we obtain six possible utterances and $6^2 = 36$ possible grammars, where a grammar G maps each of the two role-relative contexts to an utterance (e.g., $G = \{s_A > s_P \mapsto ((A, P), \emptyset), s_P > s_A \mapsto ((A, P), \sigma_P)\}$, denoting agent-first order with high-salient patient-marking).

The listener

Given context c, the listener maintains a prior over meanings based on the salience-role correlation. When $\beta_s = 2$, the listener expects the higher-salience referent to be the agent with probability $2/3$ and the patient with probability $1/3$:

$$P(m \mid c) = \begin{cases} \frac{2}{3} & \text{if } m \text{ assigns the higher-salience referent to } A \\ \frac{1}{3} & \text{otherwise} \end{cases} \tag{2}$$

The listener updates beliefs incrementally via Bayes' rule. Let u_1 denote the first sub-utterance (the first referent and any marker it carries) and u_2 the second. After observing u_1, the listener computes $P(m \mid u_1, c, G) \propto P(u_1 \mid m, c, G) \cdot P(m \mid$

[1] At Marr's computational level, scalar treatment suffices because predictions depend only on the salience–role correlation, not on which property (e.g., animacy) produces or empirically correlates with activation; decomposition is an algorithmic-level question.

[2] We exclude double-marking both for simplicity's sake (to constrain the number of possible grammars) and since such languages would have low probability anyways via production costs.

c). After observing the full utterance $u = (u_1, u_2)$, the listener computes $P(m \mid u, c, G) \propto P(u \mid m, c, G) \cdot P(m \mid c)$. Markers are semantically definitive: observing σ_A on referent r, for example, entails $r \mapsto A$ with certainty.

The speaker

The speaker chooses an utterance u to communicate meaning m in context c, balancing informativity against production costs. Production costs arise from the speaker's cognitive architecture and affect production even in the absence of any listener; they reflect the effort required to produce an utterance independent of its communicative value. The speaker's production probability follows a maximum entropy distribution subject to these competing pressures:

$$P(u \mid m, c, G) \propto I(u, m, c, G) \cdot 2^{-C(u|c)} \tag{3}$$

where $C(u \mid c)$ is the total production cost in bits. Informativity is measured as the sum of listener posteriors across time steps, rewarding early resolution of meaning:

$$I(u, m, c, G) = P_L(m \mid u_1, c, G) + P_L(m \mid u, c, G) \tag{4}$$

This formulation captures the benefit of incremental disambiguation: an utterance that resolves the listener's uncertainty after the first word is more informative than one that leaves the listener uncertain until the second word arrives.

The production cost decomposes additively into two components. The order cost reflects the sequential nature of production: both referents are simultaneously activated to some degree, and producing the less-salient referent first requires suppressing the more-salient one. This suppression incurs one bit of effort:

$$C_\omega(u \mid c) = \begin{cases} 0 & \text{if higher-salience referent is first} \\ 1 & \text{otherwise} \end{cases} \tag{5}$$

The marking cost reflects a compression bias: producing an overt marker requires additional articulatory effort beyond the unmarked case. Choosing between the two markers (σ_A or σ_P) requires specifying one bit of information, whereas zero-marking requires no specification:

$$C_\mu(u) = \begin{cases} 0 & \text{if } \mu = \emptyset \\ 1 & \text{if } \mu \neq \emptyset \end{cases} \tag{6}$$

The total production cost is the sum of these components:

$$C(u \mid c) = C_\omega(u \mid c) + C_\mu(u) \tag{7}$$

All costs were set a priori to one bit (the minimal non-trivial value), so qualitative predictions do not depend on parameter tuning. Future work could estimate these costs empirically for quantitative fits to typological frequencies.

Emergent asymmetries

Several asymmetries are apparent even without running the full model.

First, there is a marking pressure asymmetry. In prototypical contexts where $s_A > s_P$, the prior already favors the correct interpretation, so unmarked utterances remain reasonably informative. In non-prototypical contexts where $s_P > s_A$, the prior works against the speaker, creating greater pressure to mark. This follows necessarily from combining a typicality prior with a cost on marking.

Second, there is a front-loading pressure. Placing a marker on the first word yields maximal informativity ($I = 2$), while placing it on the second word delays resolution. This emerges from incremental updating without any ad hoc additional preference for early information.

Third, there is an order-marking interaction. Agent-first order with agent-marking and patient-first order with patient-marking both place the marker on the first word.

Finally, the informativity of zero-marking is grammar-dependent. While overt markers are semantically definitive, the absence of a marker is informative only to the extent that a marker is conventionalized in the grammar. If a grammar establishes that agents receive marking, an unmarked referent is more likely to be interpreted as a patient, and vice versa.

Language change and typological distribution

Grammars change as speakers occasionally deviate from convention. Because the model defines contexts in terms of referent salience (which referent is x vs. y), but grammars must generalize across specific referents, we re-express contexts in role-relative terms: c^* denotes a context defined by which thematic role is more salient (agent-salient vs. patient-salient), abstracting away from referent identity. The set $C^* = \{s_A > s_P, s_P > s_A\}$ contains the two role-relative contexts. A grammar G maps each role-relative context to an utterance, and $G(c^*)$ denotes the utterance prescribed by grammar G in context c^*.

The transition probability from grammar G to grammar G' is determined by aggregating speaker production under G:

$$T(G' \mid G) = \frac{1}{Z(G)} \prod_{c^* \in C^*} P_S(G'(c^*) \mid c^*, G) \tag{8}$$

where $Z(G)$ is a normalizing constant ensuring the transition probabilities sum to one. This captures continuous grammatical drift: speakers produce utterances according to their biases under the current grammar, and the resulting distribution determines the likely next grammar. No external prior over grammars is imposed.

The transition kernel defines a Markov chain over grammars, following the approach used in models of iterated Bayesian learning (Griffiths & Kalish, 2007; Kirby et al., 2007). The resulting 36×36 transition matrix $\mathbf{T}$, where $T_{ij} = T(G_j \mid G_i)$, characterizes the probability of transitioning between grammars. Under standard ergodicity conditions, the chain admits a unique stationary distribution π satisfying $\pi = \pi\mathbf{T}$. This stationary distribution represents the long-run typological frequencies of grammars, emerging from the interaction of speaker biases with communicative informativity.

Empirical predictions

We computed model predictions through two analyses. The static analysis examines speaker utterance probabilities under a uniform grammar prior, revealing the immediate pressures on production in each context. The dynamic analysis examines the stationary distribution over grammars, revealing which marking systems are likely to emerge and persist over time.

Static analysis Figure 1 shows the probability of each utterance type in each context. In the prototypical context (agent more salient than patient), the most probable utterance is agent-first with no marking. This reflects the alignment of multiple pressures: the production bias favors producing the higher-salience referent first, economy disfavors marking, and the typicality prior already supports the correct interpretation, reducing the need for disambiguation. The second most probable utterance is agent-first with agent-marking, followed by agent-first with patient-marking.

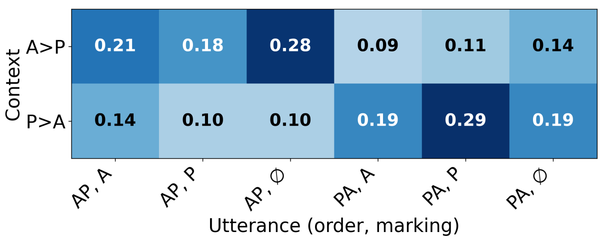

Figure 1: Speaker utterance probabilities across contexts under uniform grammar prior. Contexts refer to relative role-reference saliency (where $A > P$ is prototypical). Utterances are given as order-marking pairs. A = agent, P = patient.

In the non-prototypical context (patient more salient than agent), the pattern shifts: the most probable utterance is patient-first with patient-marking. Here the typicality prior works against the speaker: the listener expects the higher-salience referent to be the agent, but it is actually the patient. Marking resolves this conflict by providing unambiguous role information. Patient-marking on the first position is optimal because it both follows the production bias and achieves early disambiguation. The second most probable utterances are patient-first with agent-marking and patient-first with no marking, which are equally probable.

The key asymmetry is in marking pressure across contexts: in prototypical contexts, unmarked utterances remain viable because the prior supports interpretation; in non-prototypical contexts, marking becomes necessary to overcome the misleading prior, with a preference for patient marking.

Dynamic analysis Figure 2 shows the stationary distribution over grammar types by word order and marking patterns. The most striking result is that high-salience-first word order (HL) dominates: grammars producing the more salient referent first are far more probable than alternatives with uniform role-based order.

Among high-salience-first grammars, the most probable marking type is zero-marking. The second most probable is high-salient patient-marking (HP), in which only the non-prototypical context receives patient-marking while the prototypical context remains unmarked. This pattern corresponds to what the typological literature describes as typicality-based DOM: patients are marked only when they are atypical. Non-prototypical patient-marking is by far the most probable patient-marking type, substantially exceeding both uniform patient-marking (P) and prototypical patient-marking (LP).

The distribution of agent-marking types shows a strikingly different pattern. Non-prototypical agent-marking (LA) and prototypical agent-marking (HA) are approximately equally probable, and both are comparable in probability to high-salience marking (H; marking whichever referent is more salient). This lack of convergence on a single conditioning pattern contrasts sharply with patient-marking, where non-prototypical marking (HP) is strongly favored. Among grammars with uniform marking across both contexts, patient-marking (P) is more probable than agent-marking (A).

In summary, the model predicts that patient-marking systems should be more common and consistent with typicality-based conditioning than agent-marking systems, a pattern that corresponds to the empirical DOM/DSM asymmetry.

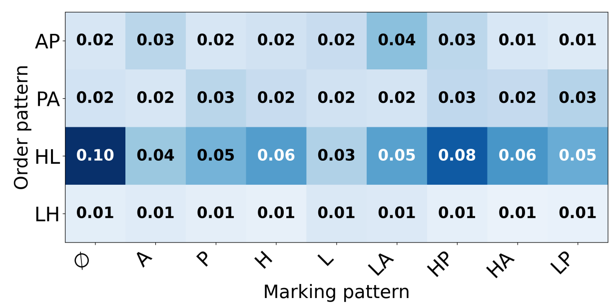

Figure 2: Long-run typological distribution of grammar types classified by word order pattern (rows) and marking pattern (columns). A = agent, P = patient, H = higher-salient referent, L = lower-salient referent.

Discussion

We now examine the sources of the predicted asymmetry, empirical validation, and limitations.

Sources of the asymmetry

The model includes a single input asymmetry: the correlation between salience and thematic role, reflecting the tendency for agents to be more salient than patients. The differential behavior of agent-marking versus patient-marking systems is not separately stipulated but emerges from how this assumption interacts with other components of the model.

If the only relevant factors were the listener's typicality prior and a speaker bias against overt morphology balanced against informativity, we would expect a symmetric distribution, where non-prototypical patient-marking and non-prototypical agent-marking are equally favored, since both disambiguate atypical configurations. Yet the predicted distribution is asymmetric. This arises from two additional factors that interact with the salience-role correlation in context-dependent ways.

First, incremental processing creates a front-loading pressure on informativity. Because the listener updates beliefs after each word, placing a marker on the first constituent resolves ambiguity earlier and yields higher informativity than placing it on the second. This pressure interacts with marking type: patient-marking on a patient-first utterance and agent-marking on an agent-first utterance both achieve maximal informativity, but which configuration is more accessible depends on which referent the speaker is biased to produce first.

Second, the activation differential between referents affects word order. Speakers find it easier to produce higher-salience referents earlier, creating a bias toward high-salience-first orders. In prototypical contexts, this bias aligns with agent-first order; in non-prototypical contexts, it favors patient-first order. The result is that uniform role-based word order is destabilized in favor of salience-based order.

This destabilization has asymmetric consequences for marking. Patient-first order in non-prototypical contexts places the patient in the position where marking is most informative, reinforcing patient-marking precisely where it is most needed. Agent-marking does not benefit from the same alignment. In prototypical contexts, the agent is both higher-salience (and therefore likely to appear first) and already expected to be the agent (reducing the benefit of marking). In non-prototypical contexts, agent-marking could aid disambiguation, but the agent is now the lower-salience referent and therefore less likely to appear first, reducing the advantage of early marking.

The interaction of these pressures produces the predicted asymmetry: patient-marking systems converge on a consistent typicality-based pattern, while agent-marking systems show no such convergence, with multiple conditioning patterns remaining comparably probable.

Empirical validation

The model generates predictions at two levels that can be empirically tested. The static analysis predicts specific production preferences in individual speakers: in non-prototypical contexts, speakers should show elevated rates of patient-marking relative to prototypical contexts. These predictions could be tested using artificial language learning experiments in which participants are trained on a language with optional case marking and then asked to communicate transitive events with a partner. Smith and Culbertson (2025) provide a methodological template, showing that differential marking emerges from communicative interaction rather than learning biases alone; however, patient-over-agent marking preferences and word order interactions remain untested.

The dynamic analysis predicts the long-run distribution of grammar types in a population. These predictions could be tested using iterated learning designs, in which the output of one generation of learners serves as the input to the next (Kirby et al., 2014; Smith, 2003). If the model is correct, iterated learning chains should converge toward high-salience-first word order with non-prototypical patient-marking, while agent-marking systems should show greater variability.

Beyond experimental tests, the model's predictions align with documented typological patterns. First, DOM should be more frequent than DSM, consistent with Malchukov (2008). Second, DOM should show consistent typicality-based conditioning while DSM should be heterogeneous. Malchukov (2008) documents precisely this: DOM cross-linguistically marks atypical objects (animate, definite, topical), whereas DSM conditioning varies substantially across languages with no dominant pattern (de Hoop & de Swart, 2007). Third, because the model predicts high-salience-first order as most probable, word order should vary with information structure, perhaps manifesting through voice alternations, fronting, or topicalization that place salient referents early regardless of semantic role (Birner & Ward, 1998; Givón, 1984; Lambrecht, 1994). Notably, Iemmolo (2020) documents that topicality and left-dislocation are cross-linguistically relevant and fairly common in DOM systems, consistent with the predicted interaction between salience-driven ordering and differential marking. These correspondences, however, require cautious interpretation. As Bickel, Witzlack-Makarevich, and Zakharko (2015) demonstrated, controlling for genealogical and areal relationships can weaken apparent universals. Rigorous tests would require phylogenetically controlled typological analysis or iterated learning experiments that can isolate these pressures from historical confounds.

Limitations

The model treats word order and case marking as the only devices for conveying thematic roles, but natural languages also employ verbal agreement, grammatical voice, and argument position relative to the verb. How these additional cues interact with case marking remains to be explored. The model defines agents and patients in purely semantic terms, whereas linguistic analyses often distinguish semantic roles from syntactic relations such as subject and object. The mapping between these levels is not straightforward, and the model does not address how syntactic structure might mediate role realization. Future work could also extend the scalar salience parameter to a vector of features (e.g., givenness, animacy), enabling predictions about interactions among referential scales.

Conclusions

Despite these limitations, the model demonstrates that an asymmetry between patient- and agent-marking systems, and the greater consistency of patient-marking with typicality-based conditioning, can emerge from the interaction of independently motivated pressures. The asymmetry need not be stipulated as a primitive of the theory; it follows from how a single input asymmetry (the salience-role correlation) interacts with incremental processing and production costs.

References

Aissen, J. (2003). Differential object marking: Iconicity vs. economy. *Natural Language & Linguistic Theory*, *21*(3), 435–483. https://doi.org/10.1023/A:1024109008573

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/s0010-0277(99)00059-1

Anderson, J. R. (1990). *The adaptive character of thought*. Lawrence Erlbaum Associates.

Bickel, B., Witzlack-Makarevich, A., Choudhary, K. K., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2015). The neurophysiology of language processing shapes the evolution of grammar: Evidence from case marking. *PLOS ONE*, *10*(8), 1–22. https://doi.org/10.1371/journal.pone.0132819

Bickel, B., Witzlack-Makarevich, A., & Zakharko, T. (2015). Typological evidence against universal effects of referential scales on case alignment. In I. Bornkessel-Schlesewsky, A. Malchukov, & M. Richards (Eds.), *Scales and hierarchies: A cross-disciplinary perspective* (pp. 7–43). De Gruyter Mouton. https://doi.org/10.5167/uzh-98911

Birner, B. J., & Ward, G. (1998). *Information status and non-canonical word order in English*. John Benjamins. https://doi.org/10.1075/slcs.40

Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, *21*(1), 47–67. https://doi.org/10.1016/0010-0277(85)90023-X

Bossong, G. (1991). Differential object marking in romance and beyond. In D. Wanner & D. A. Kibbee (Eds.), *New analyses in romance linguistics: Selected papers from the linguistic symposium on romance languages* (pp. 143–170). John Benjamins Publishing Company. https://doi.org/10.1075/cilt.69.14bos

Chater, N., & Oaksford, M. (Eds.). (2008). *The probabilistic mind: Prospects for Bayesian cognitive science*. Oxford University Press.

Cohn, N., & Paczynski, M. (2013). Prediction, events, and the advantage of Agents: The processing of semantic roles in visual narrative. *Cognitive Psychology*, *67*(3), 73–97. https://doi.org/10.1016/j.cogpsych.2013.07.002

Comrie, B. (1989). *Language universals and linguistic typology: Syntax and morphology* (2nd). University of Chicago Press.

de Hoop, H., & de Swart, P. (2007). *Differential subject marking*. Springer Netherlands.

Dixon, R. M. W. (1994). *Ergativity*. Cambridge University Press. https://doi.org/10.1017/cbo9780511611896

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505. https://doi.org/10.1111/j.1469-8986.2007.00531.x

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998. https://doi.org/10.1126/science.1218633

Givón, T. (1984). *Syntax: A functional-typological introduction* (Vol. 1). John Benjamins. https://doi.org/10.1075/z.17

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, *57*(4), 544–569. https://doi.org/10.1016/j.jml.2007.01.007

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3. speech acts* (pp. 41–58). Academic Press.

Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science*, *31*(3), 441–480. https://doi.org/10.1080/15326900701326576

Haspelmath, M. (2021). Explaining grammatical coding asymmetries: Form-frequency correspondences and predictability. *Journal of Linguistics*, *57*(3), 605–633. https://doi.org/10.1017/s0022226720000535

Holmes, J., & Vihman, V. (2025). Typicality biases in interpreting unmarked sentences: An artificial language learning experiment on differential argument marking. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *47*. https://escholarship.org/uc/item/3hc1g645

Iemmolo, G. (2020). Differential object marking. https://doi.org/10.5281/zenodo.4388380

Isasi-Isasmendi, A., Andrews, C., Flecken, M., Laka, I., Daum, M. M., Meyer, M., Bickel, B., & Sauppe, S. (2023). The agent preference in visual event apprehension. *Open Mind*, *7*, 240–282. https://doi.org/10.1162/opmi_a_00083

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62. https://doi.org/10.1016/j.cogpsych.2010.02.002

Jäger, G. (2007). Evolutionary game theory and typology: A case study. *Language*, *83*(1), 74–109.

Kirby, S., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences*, *104*(12), 5241–5245. https://doi.org/10.1073/pnas.0608222104

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, *28*, 108–114. https://doi.org/10.1016/j.conb.2014.07.014

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge University Press. https://doi.org/10.1017/CBO9780511620607

Levshina, N. (2021). Communicative efficiency and differential case marking: A reverse-engineering approach. *Linguistics Vanguard*, *7*(s3). https://doi.org/10.1515/lingvan-2019-0087

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*, 226. https://doi.org/10.3389/fpsyg.2013.00226

Malchukov, A. L. (2008). Animacy and asymmetries in differential case marking [Animacy, Argument Structure, and Argument Encoding]. *Lingua*, *118*(2), 203–221. https://doi.org/10.1016/j.lingua.2007.02.005

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. MIT Press.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, *38*(3), 283–312. https://doi.org/10.1006/jmla.1997.2543

Meir, I., Aronoff, M., Börstell, C., Hwang, S.-O., Ilkbasaran, D., Kastner, I., Lepic, R., Lifshitz Ben-Basat, A., Padden, C., & Sandler, W. (2017). The effect of being human and the basis of grammatical word order: Insights from novel communication systems and young sign languages. *Cognition*, *158*, 189–207. https://doi.org/10.1016/j.cognition.2016.10.011

New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences*, *104*(42), 16598–16603. https://doi.org/10.1073/pnas.0703913104

Scott, G. (1978). *The fore language of papua new guinea*. Department of Linguistics, Research School of Pacific Studies, Australian National University.

Smith, K. (2003). Iterated learning: A framework for the emergence of language. *Artificial Life*, *9*(4), 371–386. https://doi.org/10.1162/106454603322694825

Smith, K., & Culbertson, J. (2025). Communicative pressures shape language during communication (not learning): Evidence from case-marking in artificial languages. *Cognition*, *263*, 106164. https://doi.org/10.1016/j.cognition.2025.106164

Tal, S., Smith, K., Culbertson, J., Grossman, E., & Arnon, I. (2022). The impact of information structure on the emergence of differential object marking: An experimental study. *Cognitive science*, *46*(3), e13119. https://doi.org/10.1111/cogs.13119

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634. https://doi.org/10.1126/science.7777863