วารสารภาษาและภาษาศาสตร์ ปีที่ 39 ฉบับที่ 2 (กรกฎาคม - ธันวาคม 2564)
JOURNAL OF LANGUAGE AND LINGUISTICS 39, 2 (JULY - DECEMBER 2021)

Emergence of the Split Goal Marking System in a Population of Simulated Agents

Ponrawee Prasertsom
Attapol T. Rutherford
Faculty of Arts, Chulalongkorn University
Corresponding author's email: ponrawee.pra@gmail.com
Received 2 September 2020; revised 21 February 2021; accepted 22 February 2021; online 1 July 2021

Abstract

Languages generally prefer not to employ overt marking on motion endpoints (Goals), with this tendency being most significant for toponyms. An earlier attempt to explain this fact is incompatible with research in corpus linguistics. This paper has two aims: 1) to provide a more plausible explanation and 2) to test its validity. The explanation provided here appeals to economy: humans' desire for minimal production and comprehension effort. Its validity was tested and confirmed using agent-based modelling. The results show the contribution of each condition and its effects on the dynamics of language users' convergence on a shared marking system. Furthermore, they suggest that the current explanation may need to be amended to include another necessary condition: that language users predominantly infer the role of a zero-marked noun from overall, as opposed to verb-specific, frequency distribution of motion roles.

Keywords: linguistic typology, computational linguistics, motion event, economy

การเกิดระบบรูปบ่งชี้ปลายทางแบบจำแนกประเภทนามในแบบจำลองตัวแทนผู้ใช้ภาษา

พลรวี ประเสริฐสม
อรรถพล รัทเธอฟอร์ด
คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
Corresponding author's email: ponrawee.pra@gmail.com
รับบทความ 2 กันยายน 2563 แก้ไขบทความ 21 กุมภาพันธ์ 2564 ตอบรับ 22 กุมภาพันธ์ 2564
ออนไลน์ 1 กรกฎาคม 2564

บทคัดย่อ

ภาษามนุษย์โดยทั่วไปมักจะไม่ใช้รูปบ่งชี้ปลายทางเพื่อแสดงว่าคำนามเป็นปลายทางของการเคลื่อนที่ แต่มักจะเป็นเช่นนี้เฉพาะในกรณีที่คำนามดังกล่าวเป็นชื่อสถานที่ คำอธิบายสากลลักษณ์ข้อนี้ที่มีอยู่ขัดแย้งกับผลการศึกษาทางภาษาศาสตร์คลังข้อมูล งานศึกษาชิ้นนี้มีจุดมุ่งหมายสองประการ ได้แก่ 1) ให้คำอธิบายที่เป็นไปได้มากกว่าที่มีผู้เสนอไว้ก่อนหน้า และ 2) ทดสอบความสมเหตุสมผลของคำอธิบายดังกล่าว คำอธิบายของงานนี้อาศัยหลักความประหยัดสองประการ ได้แก่ ความต้องการของมนุษย์ในการประหยัดพลังงานในการผลิต และความประหยัดในแง่การแปลผลภาษา การทดสอบว่าสมมติฐานในคำอธิบายนี้สามารถนำไปสู่ผลที่เห็นได้จริงผ่านแบบจำลองตัวแทนผู้ใช้ภาษาในคอมพิวเตอร์ ผลจากแบบจำลองแสดงให้เห็นว่าคำอธิบายนี้สมเหตุสมผล อีกทั้งแสดงให้เห็นว่าเงื่อนไขทั้งสองข้อมีผลต่อแบบลักษณ์ของภาษาและพลวัตในการประสบผลสำเร็จในการสื่อสารอย่างไร นอกจากนี้ ผลจากแบบจำลองยังแสดงให้เห็นว่าอาจต้องเพิ่มเติมเงื่อนไขจำเป็นอีกประการหนึ่งในคำอธิบายของสากลลักษณ์ทั้งสองนี้ คือ ผู้ใช้ภาษาจะต้องคาดเดาบทบาททางความหมายของคำนามที่ใช้สูญรูปแสดงความเป็นปลายทางจากความถี่ของบทบาททางความหมายโดยรวมจากทุกกิริยาเป็นหลัก

คำสำคัญ แบบลักษณ์ภาษา ภาษาศาสตร์คอมพิวเตอร์ เหตุการณ์การเคลื่อนที่ หลักความประหยัด

1. Introduction

Stolz and colleagues (2014) found that most languages allow zero-marking on motion endpoints (Goals), expressing them without a form to indicate the Goal role. This is the case in Punjabi (Bhatia, 2010, p. 33):

tusii	Kal	bambe	Jaaoge
2PL.HON	tomorrow	Bombay	go:FUT:MASC:PL

'You are going to Bombay tomorrow.'

Out of the 112 languages in the sample, 104 have zero-marking for Goal, while only 18 for Source, the starting point of motion events.

Most languages that do employ zero-marking on Goals allow it only for toponyms (i.e., place names), but not common nouns (see Table 1).

Table 1

Distribution of Goal-marking strategies in Stolz et al.'s (2014) 112-language sample

Zero			Overt	Total
Toponyms only	Common nouns only	Both		
56 (50%)	10 (8.92%)	38 (33.9%)	8 (7.14%)	112

Given below is an example from Koiari (Dutton, 1996, p. 72):

1) negetu Era otaihere da unu
today Port Moresby go:FUT 1SG be
'I am going to Port Moresby today.'
2) mata=va Otinu
Bush=on (Goal) go.3SG.PAST
'He went into the bush.'

Stolz et al.'s (2014) findings can be summarized as two statistical universals:

1. *Preferred zero-marking on Goals:* Languages employ zero-marking on Goals more often than Sources.

2. *Preferred zero-marking on Goal toponyms:* When they do employ zero-marking on Goals, they tend to allow it only for toponyms rather than common nouns (*split Goal markin*g).

Stolz et al. (2014) have also provided explanations for these two universals. They appeal to *the goal bias*, which makes Goals cognitively salient. For instance, some psycholinguistic research has reported increased performance in change detection and event discrimination tasks involving Goals (Lakusta, 2005; Lakusta & Landau, 2012; Regier & Zheng, 2007). Stolz et al. (2014) argued that this bias makes the Goal role predictable for the addressee. Knowing this, the speaker does not explicitly mark Goals. They further proposed that zero-marking on Goal toponyms is preferred because they are always definite, so their role is easily recognized by interlocuters, making overt marking unnecessary (Stolz et al., 2014, p. 314).

Both explanations are not without problems. Their explanation for preferred zero-marking on Goals presupposes that Goals are always the most prominent motion role, but corpus analyses of English have found that this bias is not absolute: different verbs collocate with different roles (Stefanowitsch, 2018; Stefanowitsch & Rohde, 2004). For example, *escape* prefers Sources, but *go* prefers Goals. Meanwhile, the definiteness-based explanation for zero-marking on toponyms is considered implausible by even the authors themselves. They noted that in French, definite articles are obligatory with Source toponyms, yet still do not allow zero-marking.

The objectives of this study are to propose more plausible explanations for the universals and to test their validity using computational simulations, in particular *agent-based modelling*.

2. A new explanation

While problematic, Stolz et al.'s (2014) explanations provide a good basis for the new explanation being proposed here. Their explanation of zero-marking on Goals implicitly assumes the principle of *economy* (Croft, 1990; Haiman, 1983, 1985) or *least effort* (Zipf, 1935, 1949); all else equal, humans are predisposed to perform actions requiring the least amount of effort. If the Goal role is predictable, the role can be successfully conveyed to the addressee with or without overt marking, and the speaker will prefer using no form (zero-marking) to indicate Goal to minimize articulatory effort. In Stolz et al.'s (2014) explanation, the predictability of Goal rests on the unlikely assumption that the goal bias is absolute. Instead, this study adopts frequency as the source of predictability. The more frequent a meaning, the more predictable it will be (Haspelmath, 2008). While verbs prefer different roles, the most frequent overall is Goal, not Source or Trajectory (the path of motion). This is true at least in English (Stefanowitsch, 2018), Spanish and Portuguese (Becker & Naranjo, 2017). Here, this frequency-driven formal economy is termed *Articulatory Economy* (AE).

AE alone does not explain the split marking between noun types in Goals. To explain this, one can accept the insight that toponyms, marked as definite or not, have one or few possible referents, whereas common nouns have many. Mapping a toponym to its referent thus requires much less cognitive effort than doing the same for a common noun. As with articulatory effort, humans seek to minimize this sort of effort as well (e.g., Steels, 2017; van Trijp, 2017). Following Rosenbach (2002, p. 236), this economy of cognitive effort is referred to as *Cognitive Economy* (CE).

Importantly, a trade-off between AE and CE can be observed. If a noun is overtly marked, then the speaker's articulatory effort is high due presence of the overt marker, but the addressee's cognitive effort required to infer its role is low as the role can be identified with ease from the marker. If the noun is a toponym, articulatory effort is high because toponyms are generally long, but the cognitive effort is low as reference resolution is easier for toponyms. The opposite also holds in either case.

Assuming this trade-off, the current study hypothesizes that the split system arose as an optimal strategy to balance satisfactions of AE and CE. The speaker minimizes their own articulatory effort as much as possible, but at the same time avoids requiring cognitive effort on the addressee's part. The split system requires the addressee to perform only one heavy cognitive task for either noun type: reference resolution for overtly marked common nouns and frequency-based semantic-role inference for zero-marked toponyms. Zero-marked common nouns are subsequently disfavored because the addressee would be forced to perform both tasks at the same time. This "moderately selfish" behavior is likely natural since, except in rare cases such as deception, communication is cooperative (e.g., Grice, 1975), and division of effort would be expected under cooperation.

In sum, this study proposes that, given the high frequency of Goals, preferred zero-marking on Goals and preferred zero-marking on Goal toponyms (split Goal marking) emerge when the following are present:

1. Articulatory Economy (AE): Language users prefer maximally short forms among alternatives.

2. Cognitive Economy (CE): Language users prefer strategies that minimize cognitive effort.

where AE and CE apply only if communicative success is guaranteed.

This explanation does not suffer from the shortcomings of Stolz et al. (2014) because it relies on neither the absolute goal bias nor the grammatical definiteness of toponyms. Furthermore, it is not restricted to a particular linguistic theory. It is compatible with any approach that allows aspects of language use and its constraints—specifically, frequency and economy—to shape grammars. Notably, most approaches oriented around usage (Bybee & Beckner, 2015) or processing (Hawkins, 2004) would be compatible.

So far, the current explanation has only been verbally argued for. Whether it is valid—that if AE and CE are both fulfilled, the universals will follow—is unclear. *Agent-based modelling* is adopted to test the validity.

3. Methods: Agent-based modelling for hypothesis validation

Functional explanations such as the one proposed above are best tested diachronically, as only over time can one observe how the functional pressures shape a linguistic system. One prime example is a common change in which frequent forms become shorter over time due to AE and automization (Bybee & Thompson, 1997; Zipf, 1949).

However, studying language change has its limitations. Changes that help support hypotheses may not be documented. In the current case, the hypothesis regarding AE predicts a change in which a previously overtly marked role comes to allow zero-marking after that role becomes the most frequent, while overt markers become compulsory with the other roles. To the best of my knowledge, no change of this kind has been documented. Additionally, changes, even if attested, cannot clarify whether all the pressures proposed are, in fact, *required* for the emergence of the two universals. They cannot tell, for example, whether the idea that AE without CE cannot lead to a split system is correct. In the real world, CE is either relevant or is it not. The researcher cannot add or remove CE from language users at will. Nevertheless, while it is difficult to show that the explanation is empirically accurate, it is possible to show its *validity*: that the predictions made verbally will follow only when all the assumptions hold.

To do this, one could test the current explanation through simulations. Such methods are popular in studies on the origins of linguistic properties, where empirical data are available only for present-day languages (see Tamariz & Kirby, 2016 for review). Following works on the emergence of linguistic subsystems (Spranger, 2016 on spatial language; Beuls & Steels, 2013 on agreement; Pauw & Hilferty, 2012 on quantifiers), this study simulates the *emergence* of a marking system exhibiting preferred zero-marking on Goals and on Goal toponyms. This is suitable because AE and CE, inherent in humans, would have been operative since the beginning of languages; their effects would therefore be clearest when the event role marking emerges in *the first place.*[1]

Simulations equip the researcher with complete control as to which assumptions are included, how learning and interactions work and so on. This method allows the current study to construct *both* a situation in which CE is present *and* one where it is not, as well as to compare the results directly to see if CE is indeed required for split Goal marking to emerge. In addition, modelling forces one to be explicit about assumptions that are not part of the hypothesis, and as a result, the method helps reveal hidden assumptions that are required for the predictions to actually hold (Smith, 2012). This may also be the case here (see Section 6).

The current study employs *the language game model* (Steels, 2012). Under this model, agents are initially endowed with capacities, which chiefly include a language-like system lacking some features expected to develop through agents' interactions. Agents, randomly paired to communicate, attempt to understand each other, in turn, changing their grammars. The process is repeated in its entirety as many times as the researcher sets. Throughout the simulation, information on agents and their communications is recorded for analysis.

Hypothesis validation under this model proceeds as follows (Steels, 2006; van Trijp, 2017). The same model is run under different sets of assumptions, namely those with *all* the assumptions made in the hypothesis and those where one or more of the assumptions hypothesized to be required are absent. These runs of the same model under different assumptions are called

[1] Simulating change is also possible but would be unrealistic, especially when CE effects are of interest. This is due to the fact that any state of a full-fledged language would conform to CE, and a change in human cognition over an historical timescale is extremely unlikely (cf. Lass, 1997, p. 361 for an even more extreme view in which no functional pressures can trigger change.)

experiments. One concludes that the hypothesis is *valid* if an experiment that leads to a result that approximates reality is one where *all* the assumptions are present. Moreover, if a given experiment is the *only* one with such results, one may also conclude that all the assumptions are required for the hypothesis to hold.

4. Simulated data and model description

The present model is based on van Trijp's (2017) model of the evolution of case grammar. The differences between this model and van Trijp's lie mainly in the fact that the current work models only a stage of van Trijp's model (generalized semantic role markers, of which motion role markers are a subclass) and uses a different learning algorithm (4.2.3-4.2.6).

4.1 Assumptions about agents and their grammars

The current model consists of 10[2] simulated language users, or *agents.* Their only task is to repeatedly play the *description games* (van Trijp 2017, p. 52) which simulate communication about motion events potentially involving Source, Trajectory[3], and Goal. Agents are given IDs and are initially endowed with *only* the properties in 4.1.1 – 4.1.7.[4]

4.1.1 Initial lexical protolanguage with shared vocabulary

For the current purposes, data are simulated and simplified. Following van Trijp (2017), agents start out with a lexical protolanguage, a pre-language

---

[2] 10 is not a realistic number of language users in a linguistic community. The number is meant only to capture the multiplicity of language users in the real world within a degree of simplicity. Such idealizations are common for simulation studies (cf. several made in van Trijp, 2017)

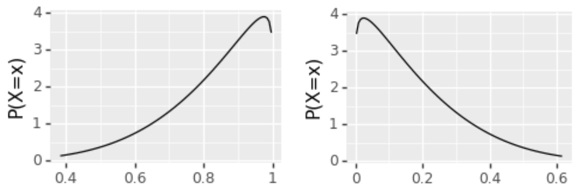
[3] Trajectory is included in the model as it is commonly associated with Source and Goal; for example, in the Source-Path-Goal schema in cognitive linguistics (Lakoff, 1987). Stolz et al.'s (2014) original claim concerns the asymmetry of Source vs. Goal and Place, the latter of which is excluded here as it belongs more properly to the domain of events in general, rather than motion events.

[4] The purpose of a simulation study is to abstract away from reality (Smith, 2012). The model includes only relevant information and assumptions. As such, demographic information of the agents (e.g., age or gender) is not given. Likewise, nouns are not specified for anything but IDs and their types. This reasoning applies for other "missing" information throughout this study.

stage with only lexical words (Bickerton, 1990), and zero markers for all three roles. Through innovation and learning (4.2), agents may later adopt overt markers used by themselves and/or other agents. Initially, all agents' inventories share a set of verbs and nouns. Verbs are all assumed to be motion verbs and come in four types: Neutral, Source-, Trajectory- and Goal-oriented, 25 verbs per type.[5] Neutral verbs have a low probability of co-occurring with any role. The others have a high probability of co-occurring with their respective orientation (Source, Trajectory, Goal) and low probabilities with the other roles. The high and low probabilities are drawn from two skewed beta distributions (Figure 1).

Figure 1

Beta distributions with α = 5, β = 1.1 (high probabilities, left) and α = 1.1, β = 5 (low probabilities, right)



4.1.2 Ability to innovate motion role markers

Agents can innovate one unique marker per role per agent. So, one agent's Source marker, for example, differs from that of the others.[6]

[5] These types are meant to represent verbs that are biased toward different roles, as found in corpus analyses (Stefanowitsch, 2018; Stefanowitsch & Rohde, 2004). In the real world, English *escape* is Source-oriented, because it occurs most frequently with Sources. In the model, these verbs are not given specific forms or meanings as these are irrelevant to the phenomenon of interest. Instead, they are only assigned probabilities of co-occurring with different roles.

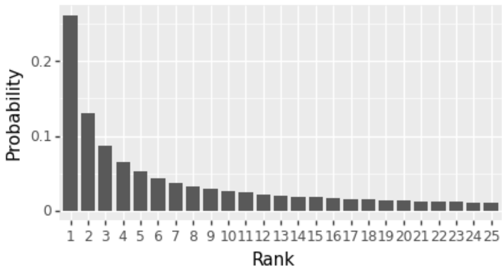
[6] In the real world, markers would come via grammaticalization and lexical diffusion. (see Lestrade, 2018 and van Trijp, 2017 for the modelling of such processes).

4.1.3 Bias to produce Goal-oriented verbs

Agents are biased toward using Goal-oriented verbs, employing these verbs 70% of the time, but only 10% of the time for each of the other types.[7] Within a particular type, the probability of agents opting for a certain verb (out of the 25) follows a Zipf distribution (Figure 2).[8]

Figure 2

Zipf distribution with a=1



4.1.4 Bias to produce common nouns

This work assumes, following real-life intuition, that common nouns are more frequent than toponyms. Agents realize event participants as common nouns 70% of the time and toponyms 30% of the time.

4.1.5 Theory of mind

Agents, like actual humans, are given the ability to reason about others' mental states (Premack & Woodruff, 1978); they can calculate the cognitive effort the addressee requires to interpret their utterances.

---

[7] This is assumed to be the source of the high frequency of Goals and is roughly compatible with Stefanowitsch's (2018) results from English corpora.

[8] This choice of distribution serves only to build into the model the fact that 1) the role a given verb is biased toward is more common than others (beta) and 2) some verbs are more frequent than others (Zipf). They do not constitute claims about actual motion verb distributions.

4.1.6 Full cooperation

Agents fully cooperate in communication.[9]

4.1.7 Bias for conformity

Agents use other agents' markers if already available, rather than trying to spread their own innovations (following van Trijp, 2017).

4.2 The description game

An *iteration* is a series of description games played by each of the 10 agents in a random order. A single *run* of an experiment consists of 10,000 iterations.[10] Thus, there are 100,000 games per *run*. Each *description game* simulates communication about motion events between two agents. A single description *game* is comprised of the steps given in 4.2.1 - 4.2.6.

4.2.1 Role assignment

The agent of that game is assigned the speaker role. Another agent is randomly selected and assigned the addressee role.

4.2.2 Event generation

The verb is selected according to the distributions specified in 4.1.3. Nouns in different roles are generated according to the probabilities given by the verb. An event consists of the verb and nouns, coded as an *event vector* of 9 dimensions: 3 dummy-coding whether the event has each role (Source, Trajectory, Goal), 3 dummy-coding whether each noun is a common noun and 3 coding whether each noun is a toponym.[11]

4.2.3 Production

For each role, the speaker agent selects either to zero-mark or overtly mark using a marker from its inventory. This selection is determined by scores obtained by an algorithm based on the perceptron (Rosenblatt, 1958). The

[9] This is an idealizing assumption made to simplify the process since lying is relatively rare in real life and of little relevance to the current model.

[10] This is set so that agents play enough games to allow them to learn and align their marking strategies with other agents and eventually come to share the same marking system.

[11] For example, an event with only a Goal toponym is coded as [0 0 1 0 0 0 0 0 1], with the 3rd and 9th dimensions showing that there is a Goal and that it is a toponym, respectively.

algorithm is as follows. First, all possible strategy combinations (zero/overt markers) are generated according to the agent's inventory and dummy-coded in *strategy vectors* of 33 dimensions. Three dimensions code whether zero-marking is used for each role, and 3 roles × 10 agents = 30 dimensions code which overt markers are used. Then, *cross vectors* with 2 roles × 33 strategies = 66 dimensions are generated. Each dimension dummy-codes an interaction of an event role and a strategy. The first, for example, codes the presence of a Source common noun. The event, strategy and cross vectors are concatenated into the event-strategy-cross (ESC) vector.

Each agent has a corresponding 109-dimensional *weight* vector used to calculate scores, with 9 dimensions for the event vector, 33 for the strategy vector, 66 for the cross vector and 1 bias dimension. The dot product[12] between the ESC vector of a strategy (with 1 appended at the end as the bias term) and the weight vector is the score for that strategy. If the experiment assumes AE and/or CE, the score is subtracted by the AE and/or CE cost. The AE cost is operationalized as

$$\frac{(\text{No. of toponyms } + \text{No. of markers})}{2 \times \text{No. of roles}}.$$

In effect, it is highest (= 1) when all nouns in the event are both toponymic and overtly marked and lowest (= 0) when they are neither. The CE cost formula is $CE = 1 - \textbf{AE cost},$ reflecting the AE-CE trade-off. The final score of each strategy is obtained by applying the sigmoid function to the cost-subtracte score, thereby scaling it down to between 0 and 1. The strategy with the highest score is selected and its ESC vector is treated as the "utterance" of the game.

4.2.4 Comprehension

The addressee decodes the ESC vector to obtain all event and noun information, except nouns' roles. Known markers are mapped to the correct

---

[12] A dot product of two vectors and is $a = [a_1 \quad a_2 \quad a_3 \quad \ldots \quad a_n]$ and $b = [b_1 \quad b_2 \quad b_3 \quad \ldots \quad b_n]$ is $\sum_{i=1}^{n} a_i b_i$.

roles. If a noun is zero-marked or marked with an unknown marker, the role is inferred based on the overall frequencies of role combinations. For instance, if an utterance has a noun with a Source marker and a zero-marked noun, the addressee will find the most frequent *pair* of roles that includes a Source in its experience of all two-role events. If the verb is among the 10 most frequently encountered by the agent, however, the guess will employ verb-specific frequency statistics instead.[13] For three-participant events, agents can guess the remaining role if two nouns are marked with known markers.

4.2.5 Feedback

Both agents learn whether the roles and nouns are correctly mapped. If they are, the communication is considered successful.[14]

4.2.6 Weight and inventory updates.

The addressee learns any marker in the utterance unknown to it. Weights are updated according to the equation $New weight_i = Old\ weight_i + CS \times \alpha \times x_i \times (1 - Score)$, where α is the *learning rate*, set at 0.01[15]; is the dummy variable in the ith dimension; Score is the score of the strategy used; and CS (communicative success) is 1 when the communication is successful and is -1 otherwise.

Weight updating is the key mechanism by which agents learn to prefer one strategy over others. If a strategy succeeds more often, the weights associated with that strategy will be increased more throughout the games. This in turn increases its strategy score, and hence preference, in the Production step.

The agents' role combination frequency statistics, overall and verb-specific, are updated. Agents also apply a conformity bias every 10 games by removing the least used marker within the 10-game period from the

[13] This decision is based on finding that language users utilize specific statistics for high-frequency verbs (Wonnacott, 2011; Wonnacott et al., 2008; see also Bybee, 2013 for the idea that high-frequency tokens are autonomous and do not activate the type cluster of the verb).

[14] This is assumed to be extralinguistic cues of understanding in actual interactions.

[15] The learning rate dictates the amount of weight change per update. Pilot runs indicate that 0.01 is small enough to allow time for the agent to learn and come to share preferred strategies.

inventory of each role unless only one overt marker is left for that role.[16]

4.3 Implementation

The model was implemented in Python, using the MESA package (Masad & Kazil, 2015). Scipy (Virtanen et al., 2020) and Pandas (The Pandas development team, 2020) were used for data generation and analysis.

5. Experiments

Four experiments were conducted, with the first three lacking one or both assumptions in the hypothesis. Since the hypothesis requires both AE and CE in order for the Goal-only split marking system to emerge, only Experiment 4, where both AE and CE are present, is predicted to result in preferred zero-marking on Goals and Goal toponyms, whereas the others are predicted to yield a different result (see 5.1-5.4 for full reasoning).

Table 2

Comparison of the four experiments. AE = Articulatory Economy; CE = Cognitive Economy

Experiment	Conditions		Predicted preferred strategy of agents	
	AE	CE	Source & Trajectory	Goal
1			Overt only	Overt & Zero (random)
2	✓		Overt only	Zero only
3		✓	Overt only	Overt only
4	✓	✓	Overt only	Overt for common nouns Zero for toponyms

[16] This process is adapted from memory decay (van Trijp, 2017), which helps agents converge on the same set of markers.

Two measures are examined and reported below: *(global) communicative success* and *marking preference.* In each game, the (local) communicative success is 1 if the addressee correctly maps nouns to their roles, 0 otherwise (cf. 4.2.5). The reported communicative success is *global*, obtained by calculating the moving average of the local measure over 1,000 games. To measure marking preference, agents are fed *test events*—all 27[17] possible event types for strategy selection. Marking preference is the number of agents that prefer overt/zero-marking for each role-noun type combination in the test events, averaged across all 10 runs of each experiment. Communicative success indicates whether the system is still functional (a crucial property for a realistic system), whereas marking preference is used to (dis)confirm the predictions.

5.1 Experiment 1 (baseline)

In this experiment, no additional assumptions are made besides those indicated in Section 4.1. It is predicted that:

1) Agents will *converge* on (come to use) the same overt markers as they will have been exposed to each other's markers and align strategies through conformity bias (see 4.1.7).

2) The system will be functional; communicative success will be close to 1 by the end because agents will have preferred the same markers.

3) Split marking dependent on noun types will not occur because there is neither AE nor CE to drive the split.

4) Agents will prefer overt markers for Source and Trajectory because they are better than zero markers when there is no AE or CE, requiring only that the marker is known and no guesses have to be made.

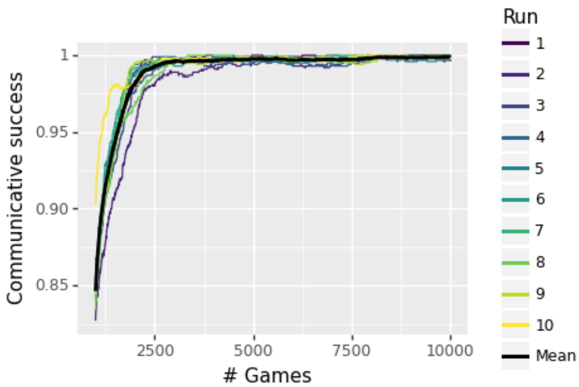
[17] An event type is a combination of roles and noun types. Roles in events are assumed to be unique (see 4.2.2), so a combinatoric calculation yields: 1 zero-role event (no nouns) + 3 roles x 2 noun types = 6 one-role events (Source-common noun, Source-toponym, Trajectory-common noun, ...) + 12 two-role events (Source-common noun + Trajectory-common noun, Source-common noun + Trajectory-toponym, ...) + 8 three-role events = 27 possible event types.

5) Agents' Goal marking preference will split roughly equally between zero- and overt marking. This is because zero-marked Source or Trajectory nouns will be wrongly interpreted as Goal, the most frequent role. For Goals, either overt or zero-marking can achieve high communicative success. Because inference from frequency will almost always be correct, it should be up to chance whether agents will have higher weight values for overt or zero-marking at the end.

Results show that communicative success increased and stabilized after about 2,500 games (Figure 3), eventually fluctuating around 0.999, thereby suggesting the system was functional.

Figure 3

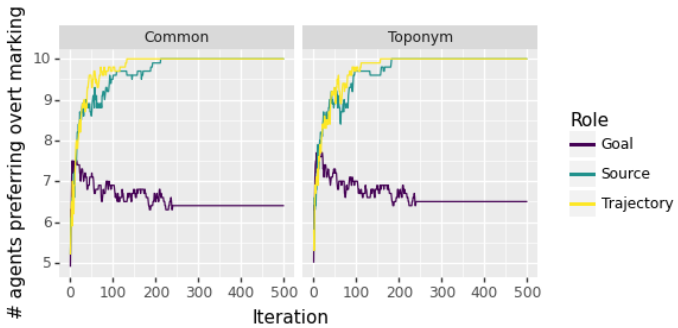
Average communicative success in the first 10,000 games in Expt. 1



All predictions are confirmed. The only deviation from the predictions is a slight bias toward overt markers for Goals. Thirty-six out of 100 agents in the 10 runs preferred zero markers for common noun Goals. Those same agents also showed the same preference for toponym Goals, except one agent that exhibited split Goal marking. Figure 4 illustrates the dynamics of marking preference across iterations.

Figure 4

Average no. of agents choosing to use overt markers with most test events involving each role as common nouns and toponyms over the first 500 iterations (5,000 games) in Expt. 1



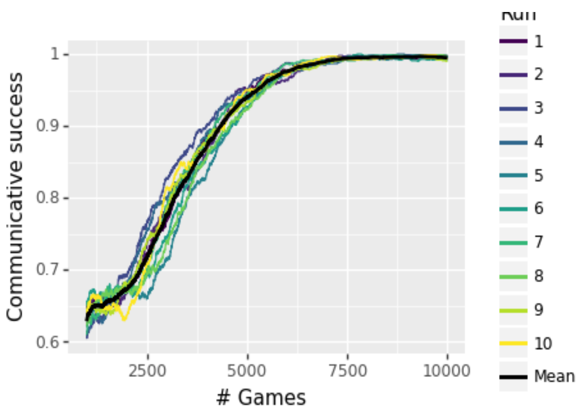
5.2 Experiment 2 (AE)

The second experiment adds an AE constraint on top of the first. The predictions are the same as Experiment 1, except that zero-marking is now predicted to apply to Goals regardless of noun type. This extends from the fact that it performs communicatively as well as overt marking, but its scores are not penalized by AE costs like overt marking.

It was found that agents reached communicative success of about 0.997 after 7500 games (Figure 5).

Figure 5

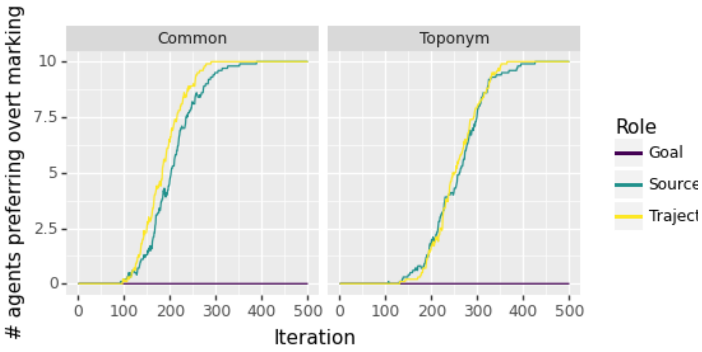
Average communicative success in the first 10,000 games in Expt. 2



The predictions are confirmed by the fact that all agents converged on overt marking within the first 500 iterations (5,000 games), while for Goals, agents opted for zero-marking throughout (Figure 6).

Figure 6

Average no. of agents with preferred overt marking within the first 500 iterations (5,000 games) of Expt. 2

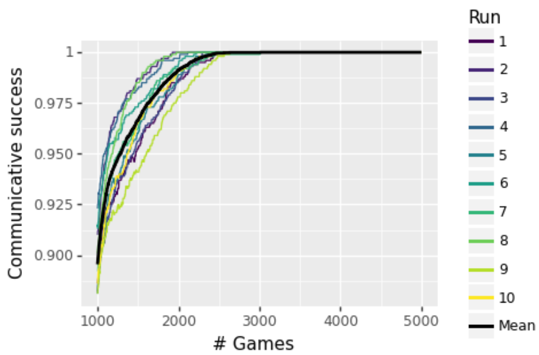


5.3 Experiment 3 (CE)

This experiment is likewise built on top of Experiment 1 by adding CE but not AE. It is predicted that overt markers will be preferred for all roles and noun types because cognitive effort is minimized when overt marking is maximized. Using overt markers will achieve higher scores even if zero-marking performs equally well as the latter violates CE.

Figure 7

Communicative success over the first 5,000 games in Expt. 3



In all 10 runs, the agents quickly reached a communicative success of 1 within 3,000 games (300 iterations). As predicted, at the end, agents preferred overt marking regardless of roles and noun types.[18]

5.4 Experiment 4 (AE+CE)

This experiment incorporated both AE and CE. If AE and CE costs are simply combined by addition, the total cost will always be 1 because the AE-CE tradeoff is assumed $(\text{CE} = 1 - \text{AE cost})$. In this case, the results will be no different from having no constraints (as in Experiment 1) as the scores are simply always shifted down by 1.

Recall, however, that the hypothesis does not state that articulatory and cognitive effort are minimized *togetyher* as this would be meaningless given the trade-off. Instead, it states that speakers minimize articulatory effort *only up to the point* where cognitive cost is not too high for the addressee. As such, unlike Experiment 3, here CE is operationalized as a threshold. When the CE cost exceeds 0.5, the strategy is dramatically disfavored (the vector score is subtracted by 999,999,999).

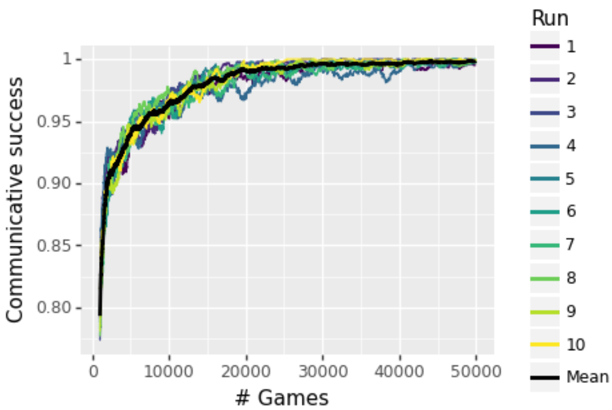
[18] Plots for preferred strategies are omitted for the sake of brevity. The number of agents with preferred overt marking remains at 10 in all games, resulting in overlapping straight lines at 10.

The prediction here is that the results will approximate reality. Split marking should be observed in the Goal domain, while for the other two roles, overt marking should be preferred across noun types.

As Figure 8 shows, communicative success increased much more slowly than in the previous experiments but became stable after about 25,000 games, oscillating around 0.998 by the end.

Figure 8

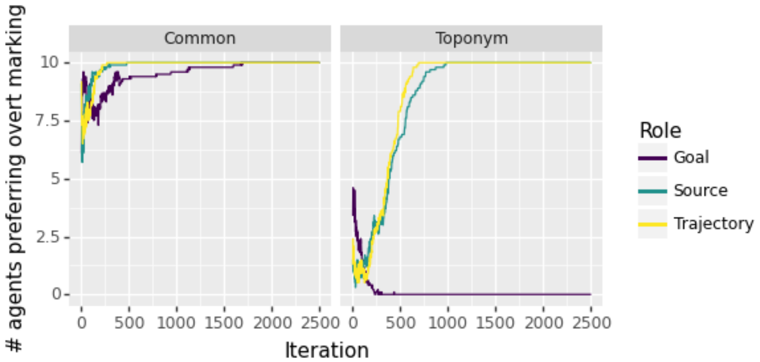
Communicative success in the first 50,000 games in Expt. 4



The predictions are again confirmed. Across runs, all agents came to prefer zero-marking for Goal toponyms and overt marking for Goal common nouns within the first 2000 iterations (Figure 9). A small exception is the two agents in the 3rd run, which still preferred zero-marking for common noun Goals for two test events at the end.

Figure 9

Average no. of agents with preferred overt marking within the first 2500 iterations (25,000 games) of Expt. 4



6. Discussion

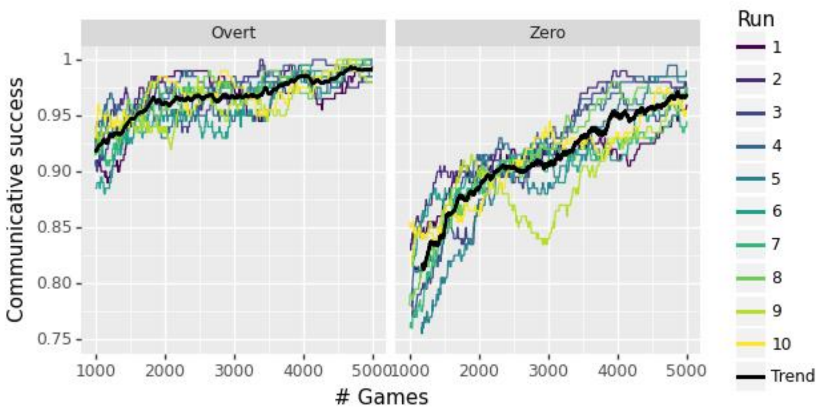
Recall that the hypothesis here stated that AE and CE are both required for preferred zero-marking on Goal and Goal toponyms to emerge. The experiments validated this hypothesis. As predicted, only in Experiment 4 (AE+CE) can one observe both preferred marking on Goals and preferred marking on Goal toponyms. The resulting system was functional: the communicative success was close to 1. Chance exceptions occurred but in general do not affect the explanation. Apart from these conclusions, the results yield further insights.

First, the unexpected slight bias toward overt markers for Goals in Experiment 1 sheds light on the behavior of language users if AE and CE were to not exist. The bias can be explained by the larger number of overt markers (10 for each role) compared to zero markers (1 for each role). In the early games, agents rapidly adopted new markers, and frequency differentials among roles were still not accurate enough for agents to infer roles correctly without overt marking.

Moreover, even if one overt marker fails, there are other overt markers for agents to choose from, while zero is the sole strategy of its kind. This resulted in lower communicative success for zeros (Figure 10). This bias was offset in certain cases where agents came to prefer zeros when the preferred overt marker was not used by other agents. Still, the overall result was a weak bias toward overt marking.[19]

Figure 10

Communicative success (moving average over 200 games) of events in Expt. 1 games 1000-5000, categorized by Goal marking strategy. The trend is calculated by moving-averaging the mean of runs in Expt. 1 every 50 games. The lines automatically connect where the event does not have a Goal.



Another noticeable result is that Experiment 2 was the only experiment in which agents achieved perfect communicative success, and it saw the shortest time span to high communicative success. The perfect success is thanks to the fact that, once they had learned the overt markers used by others, addressees *always* correctly mapped nouns to their roles. The high speed of success increase was driven by the fact that, once known to the speaker and addressee, overt markers never failed; they only failed the first time an agent encountered them during early games.

[19] In the real world, the bias should exist as well, but could be even weaker. Actual innovations would coincide more often than in the model because sources of grammaticalization tend to be limited and recurrent across languages (Kuteva et al., 2019).

In the other experiments, rare communicative failures occurred when speaker agents zero-marked Goals of an event with one of the 10 most frequent non-Goal-oriented verbs. The addressee then used the verb-specific statistics, leading them to incorrectly map the zero-marked noun to a role other than Goal. One such game is given in Table 3.

Table 3

Example of post-convergence games where communication fails (Expt. 2, 10th run, Game #99675)

Speaker's event & strategy				Addressee's inference		Success?
Noun ID	Roles	Noun type	Marking	Roles	Stats used	No
56	Goal	Toponym	Zero	Trajectory	Verb-specific	
97	Source	Common	Overt	Source		

Generally speaking, failures like this happened because zeros can be freely interpreted as any motion role with the highest overall/verb-specific frequency in the model. If zeros were conventionalized to refer to Goals only, these rare cases would disappear. Yet, it is not clear whether Goal-marking zeros are conventionalized this way in the real world or not. The existence of cases like Table 3 in real languages would imply that languages may tolerate occasional communicative failure for the sake of economy. To my knowledge, there are no detailed statistics on verbs and their co-occurrence with motion roles across languages with zero Goal marking, so this case remains to be explored in future studies.

Finally, the model has highlighted the possibility of another condition for a marking system with the Goal-only split. In the model, the agents predicted the role of zero-marked nouns based on overall frequency distribution of motion roles, except for high frequency verbs. Impressionistically, if agents had relied on individual statistics for every verb, then the split marking should have figured in all roles as these roles would have been predictable.

Hence, the use of overall statistics seems to be a necessary condition for asymmetry between Goals and the other two roles, in addition to AE and the known fact that Goals are high in frequency. This merits future investigations as, if it turns out to be necessary, the current hypothesis includes inference from overall statistics as another condition. This insight demonstrates well the advantage of modelling noted by Smith (2012): this condition has been overlooked when the hypothesis was formed verbally, but modelling helped reveal this hidden assumption.

Despite its validity and the insights gained, the model explains only 50% of Stolz et al.'s (2014) languages, which conform to both preferred zero-marking on Goals and preferred zero-marking on Goal toponyms. Why do 33.9% of the languages always zero-mark Goals? Why do 8.92% allow zero-marking of common nouns only, while 7.14% always require overt marking for Goals? There are undoubtedly relevant factors beyond AE and CE that affect language types, and surprising cross-linguistic differences might exist. For example, common nouns in some languages might be longer than toponyms, hence common-noun-only zero-marking. These are only speculations, however, and more research should be conducted to answer such questions adequately.

7. Acknowledgements

An updated and extended version of this study is part of the first author's master's thesis at the Department of Linguistics, Faculty of Arts, Chulalongkorn University: *Modelling the Evolution of Linguistic Goal Marking.* The first author would like to thank Naruadol Chancharu for comments and advice at various stages. Thanks also go to Ashley Laughlin and Daniel Loss for proofreading the manuscript. Any remaining errors are our own.

References

Becker, L., & Naranjo, M.G. (2017). Coding asymmetries, frequency and predictability: The case of to vs from [Conference presentation abstract]. *The Annual Conference of the German Linguistic Society, Germany, 39*, 144-146. http://dgfs2017.unisaarland.de/wordpress/abstracts/AG5/ag5-2-beck.pdf.

Beuls, K., & Steels, L. (2013). Agent-based models of strategies for the emergence and evolution of grammatical agreement. *PLoS ONE, 8*(3).

Bhatia, T. K. (2010). *Punjabi: A cognitive-descriptive grammar.* Routledge.

Bickerton, D. (1990). *Language and species.* Chicago University Press.

Bybee, J. L. (2013). Usage-based theory and exemplar representation of constructions. In T. Hoffman & G. Trousdale (Eds.), *The Oxford handbook of construction grammar* (pp. 46-69). Oxford University Press.

Bybee, J, L., & Beckner, C. (2015). Usage-based theory. In B. Heine & H. Narrag (Eds.), *The Oxford handbook of linguistic analysis* (2nd ed., pp. 953-980). Oxford University Press.

Bybee, J., & Thompson, S. (1997). Three frequency effects in syntax. In M. L. Luge & J. L. Moxley (Eds.), *Proceedings of the Annual Meeting of the Berkeley Linguistics Society, 23,* 378-388. https://doi.org/10.3765/bls.

Croft, W. (1990). *Typology and universals.* Cambridge University Press.

Dutton, T. E. (1996). *Koiari.* Lincom Europa.

Grice, P. (1975). Logic and conversation. *Studies in the way of words.* Harvard University Press.

Haiman, J. (1983). Iconic and economic motivation. *Language, 59*(4), 781-819.

Haiman, J. (1985). *Natural syntax.* Cambridge University Press.

Haspelmath, M. (2008). Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics, 19*(1), 1-33.

Hawkins, J. A. (2004). *Efficiency and complexity in grammars.* Oxford University Press.

Kuteva, T., Heine, B., Hong, B., Long, H., Narrog, H., & Rhee, S. (2019). *World lexicon of grammaticalization* (2nd ed.). Cambridge University Press.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* The University of Chicago Press.

Lakusta, L. (2005). *Source and goal asymmetry in non-linguistic motion event representations* [Unpublished doctoral dissertation]. The John Hopkins University.

Lakusta, L., & Landau. B. (2012). Language and memory for motion events: Origins of the asymmetry between source and goal paths. *Cognitive Science, 36*(3), 517-544.

Lass, R. (1997). *Historical linguistics and language change.* Cambridge University Press.

Lestrade, S. (2018). The emergence of differential case marking. In I. A. Seržant & A. Witzlack-Makarevich (Eds.), *Diachrony of differential argument marking* (pp. 481-508). Language Science Press.

Masad, D., & Kazil, J. (2015). Mesa: An agent-based modeling framework. *Proceedings of the Python in Science Conference, 14,* 51-58. http://doi.org/10.25080/Majora-7b98e3ed-009

Pauw, S., & Hilferty, J. (2012). The emergence of quantifiers. In L. Steels (Ed.), *Experiments in cultural language evolution* (pp. 277-304). John Benjamins.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and brain sciences, 1*(4), 515-526.

Regier, T., & Zheng, M. (2007). Attention to endpoints: A cross-linguistic constraint on spatial meaning. *Cognitive Science, 31*(4), 705-719.

Rosenbach, A. (2002). Genitive variation in English: Conceptual factors in synchronic and diachronic studies. Mouton de Gruyter.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review, 65*(6), 386-408.

Smith, K. (2012). Why formal models are useful for evolutionary linguists. In M. Tallerman & K. R. Gibson (Eds.), *Oxford handbook of language evolution* (pp. 581-588). Oxford University Press.

Spranger, M. (2016). *The evolution of grounded spatial language.* Language Science Press.

Steels, L. (2006). How to do experiments in artificial language evolution and why. In A. Cangelosi, A. Smith, & K. Smith, (Eds.), *Proceedings of the 6th International Conference on the Evolution of Language.* World Scientific Publishing.

Steels, L. (2012). Introduction: Self-organization and selection in cultural language evolution. In L. Steels (Ed.), *Experiments in cultural language evolution* (pp. 1-37). John Benjamins.

Steels, L. (2017). Human language is a culturally evolving system. *Psychonomic bulletin & review, 24*(1), 190-193.

Stefanowitsch, A. (2018). The goal bias revisited: A collostructional approach. *Yearbook of the German Cognitive Linguistics Association, 6*(1), 143-166.

Stefanowitsch, A., & Rohde, A. (2004). The goal bias in the encoding of motion events. In G. Radden & K. Panther (Eds.), *Studies in linguistic motivation* (pp. 249-268). Walter de Gruyter.

Stolz, T., Lestrade, S., & Stolz, C. (2014). *The crosslinguistics of zero-marking of spatial relations.* De Gruyter.

Tamariz, M., & Kirby, S. (2016). The cultural evolution of language. *Current Opinions in Psychology, 8,* 37-43.

The Pandas Development Team. (2020). Pandas 1.1.4. Zenedo. https://doi.org/10.5281/zenodo.4161697

van Trijp, R. (2017). *The evolution of case grammar.* Language Science Press.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Walt, S., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A., Jones, E., Kern, R., Larson, E., ... Scipy 1.0 Contributors. (2020). Scipy 1.0: Fundamental algorithms for scientific computing on Python. *Nature Methods, 17*(3), 261-272.

Wonnacott, E. (2011). Balancing generalization and lexical conservatism: An artificial language study with child learners. *Journal of Memory and Language, 65*(1), 1-14.

Wonnacott, E., Newport, E. L., & Tenenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology, 56*(3), 165-209.

Zipf, G. K. (1935). *The psycho-biology of language.* Houghton-Mifflin.

Zipf, G. K. (1949). *Human behavior and the principle of least effort.* Addison-Wesley Press.