Emergence of differential object marking in Asia Minor Greek

A computational approach to language change

Ümit Atlamaz and Metin Bagriacik
Boğaziçi University

This paper investigates the emergence of differential object marking (DOM) in the Asia Minor Greek dialect of Pharasa (PhG) under contact with Turkish. We show that DOM in Turkish and PhG are both instances of structural accusative case and DOM can be formally modeled as context sensitive dependent case. We propose that two factors caused the emergence of DOM in PhG, namely (i) case neutralization in indefinite contexts, and (ii) an increase in the number of V-NP idioms borrowed from Turkish where the NP is in bare form. These perturbations led to a significant change in the overall data created by the community resulting in mixed input for the younger generations. Once the amount of bare NPs passed a certain threshold, a divergent grammar became inevitable. We test our proposal using an abductive generalization learning algorithm based on the Tolerance Principle and running a number of simulations. Our simulation results confirm our hypothesis.

Keywords: language contact, language change, mixed input, differential object marking, Tolerance Principle, abductive learning, Asia Minor Greek, Turkish

1. Introduction

We re-examine Differential Object Marking (DOM) in the Asia Minor Greek dialect of Pharasa (hereafter PhG) and investigate how DOM might have emerged as a function of learning from mixed input.[1] Specifically, we (a) provide a detailed

1. The discussion in this paper would probably carry over to another Asia Minor Greek dialect, Cappadocian Greek, which also exhibits DOM. For reasons of space however we limit the empirical focus of this paper to PhG.

https://doi.org/10.1075/lv.23019.atl | Published online: 17 May 2024
Linguistic Variation ISSN 2211-6834 | E-ISSN 2211-6842

formal analysis of DOM in PhG and Turkish, and (b) propose an acquisition based language change model that accounts for the emergence of DOM in PhG, subscribing to the by now unanimously accepted claim that the emergence of DOM in PhG should be associated with contact with Turkish (Dawkins 1916, pp. 94, 165; Janse 2004; Spyropoulos and Tiliopoulou 2006; Karatsareas 2011, 2020; Spyropoulos 2020), as opposed to analogical extension (Andriotis 1948) or inheritance (Anastasiadis 1976).

Analyzing some core properties of DOM in Turkish and PhG, we argue that a feature relativized version of the Dependent Case Theory (Marantz 1991; Baker 2015) accounts for the distribution of DOM in both languages. We define DOM as the output of a decision mechanism that determines whether the downward dependent case (i.e. accusative) applies to a noun phrase in a given context, where the relevant context is determined by the presence of a particular feature (e.g. specificity, definiteness, etc.). In this model, a language will mark its internal arguments differentially if the downward dependent case is associated with a particular feature and it will mark them consistently (non-differentially) if no such association is present. We define the task of acquiring a case grammar that produces DOM as learning the contextual features associated with the downward dependent case.

Following Andersen (1973), Lightfoot (1979; 1991), Roberts (2021) and Kodner (2020) among others, we argue that the acquisition process can lead to learner grammars that are distinct from baseline grammars whose output is the input for the learner, especially when the input is ambiguous or substantially different from the one that was available to the baseline speakers. To that end, we claim that the main driving factors behind the emergence of DOM in PhG were loss of clues for case distinction in indefinite contexts and the increased number of V-NP fixed expressions borrowed from Turkish, where the NPs are in bare form. We argue that these V-NP fixed expressions were calqued over an extended period of time by different members of the community whose production rules may or may not have been impacted. However, when the amount of such expressions passed a certain threshold, the change in the total input inevitably led to a reanalysis and divergence in the grammar acquired by the children resulting in DOM. To test our hypothesis, we use an abductive generalization learning algorithm by Belth et al. (2021) which is based on the Tolerance Principle (Yang 2016) and run a number of simulations on Turkish and PhG patterns. Our results show that the algorithm can learn DOM in Turkish successfully. Further, we show that the model learns PhG-type DOM when the type frequency of the bare NPs reaches a certain threshold, supporting our claim that an increased number of V-NP calques might indeed be the ultimate culprit.

The organization of the paper is as follows: Section 2 introduces DOM in PhG. Section 3 compares DOM in PhG with Turkish to establish the core properties of the two systems. Section 4 provides a formal account of the DOM phenomena in the two languages within the tenets of the Dependent Case Theory. Section 5 discusses the emergence of DOM in PhG as a corollary of the language acquisition process where the input is altered through calqueing from Turkish. It further presents the details of the learning algorithm, simulations, and results. Section 6 concludes the discussion.

2. DOM in the Greek dialect of Pharasa

PhG belongs to the inner Asia Minor Greek dialect group together with the varieties of Silli, Cappadocia, and Pontus (Dawkins 1916, pp. 205–206, 1940, p. 23; Triantafyllidis 1938, pp. 273–295; Karatsareas 2011, p. 40). At the beginning of the twentieth century, it was spoken by about 2600 people in six villages around modern-day Kayseri in central Turkey (Sarantidis 1899, p. 121; Dawkins 1916, pp. 32–35). Between 1923 and 1925, its speakers were relocated to Greece in the context of the population exchange between Greece and Turkey, enacted as a supplementary protocol to the Treaty of Lausanne. Today, PhG is a moribund variety spoken by around 25 people in a number of villages in northern Greece. As of 2020, all its speakers were second generation refugees, their ages ranging between 66 and 84. The empirical domain, namely DOM, is observed both in the available texts written before the population exchange and in the recordings obtained from current speakers. Given the overlapping DOM facts between the two sources, we take the native speakers we have consulted as acrolectal speakers, whose grammar is a good approximation of what one can derive from the older texts, and receive acceptability judgments from them when necessary.

Internal arguments in PhG, both as spoken today and as observed in the older texts, must bear accusative case (1) if they are definite, and appear in nominative when indefinite (2) (Levidis 1892; Dawkins 1916; Favis 1948; Andriotis 1948; Anastasiadis 1976; Janse 2004; Karatsareas 2011).

(1) a. To čočuxi faiz ton tana-Ø.
the.NOM child.NOM feed.3SG the.M.ACC.SG calf.M-ACC.SG
'The child is feeding the calf.'

b. *To čočuxi faiz ton tana-s.
the.NOM child.NOM feed.3SG the.M.ACC.SG Calf.M-NOM.SG

(2) a. To jadhi pičin (an) tana-s.
the.NOM cow.NOM made.3SG (an) calf.M-NOM.SG
'The cow made a calf.'

b. *To jadhi pičin (an) tana-Ø.
the.NOM cow.NOM made.3SG (an) calf.M-ACC.SG

The accusative-nominative distinction appears only on masculine singular stems (cf. highlighted in gray in Table 1). The case markers are syncretic on singular feminine and neuter stems as well as on all nouns in the plural. This means DOM is visible only on masculine singular stems (Karatsareas 2011).

Case distinction is also visible on the masculine and feminine definite articles (singular or plural) but not on the neuter ones (Table 2). The indefinite article is also invariant; *an* 'a(n)'.

Table 1. Accusative-nominative case on nouns in PhG

		MASCULINE			FEMININE	NEUTER
SG	NOM	apo-s	nomati-s	prakana-s	neka	praði
	ACC	apo-Ø	nomati-Ø	prakana-Ø	neka	praði
PL	NOM	api	nomati	prakanaðe	necis	praðe
	ACC	api	nomati	prakanaðe	necis	praðe
		'fox'	'man'	'insect'	'woman'	'foot'

Table 2. Definite articles in PhG

	MASCULINE		FEMININE		NEUTER	
	SG	PL	SG	PL	SG	PL
NOM	o	i	i	i	to	ta
ACC	ton	tis	tin	tis	to	ta

3. A comparison of DOM in Turkish and PhG Greek

This section compares DOM in Turkish and PhG along two dimensions consisting of the features involved in DOM to the structural configurations in which DOM is observed.

3.1 Features involved in DOM

Cross-linguistically, DOM correlates with a set of discourse related features such as definiteness, specificity, animacy, etc. Aissen (2003) noted that an object is more likely to be differentially marked if it is on the higher end of discourse prominence hierarchies like the *Animacy Hierarchy* and the *Definiteness Hierarchy* (3), proposed by Silverstein (1976), Comrie (1979), and Croft (1988), among others.

(3) *Definiteness Hierarchy*
Pronoun > Name > Definite > Indef. specific > Non-specific

In both PhG and Turkish, DOM is associated with the *Definiteness Hierarchy* and is identificational (Kornfilt 2009; Spyropoulos 2020) namely DOM on an NP (and lack thereof) correlates with a specific property in (3) that the NP bears. Furthermore, given the implicational nature of (3), any higher feature than the one typically associated with DOM also triggers DOM.

The cut-off point for DOM in PhG is *definiteness* (Dawkins 1916; Karatsareas 2011; Spyropoulos 2020): definite full NPs (cf. (1)), proper nouns (4), and (weak and strong) pronouns (5) consistently receive accusative case when they are internal arguments.

(4) Idha ton Murati-Ø / *Murati-s.
saw.1SG the.ACC Murat-ACC / Murat-NOM
'I saw Murat.'

(5) Idha se/sena / *si.
saw.1SG you.cl.ACC/you.ACC / you.NOM
'I saw you.'

Regardless of their specificity, indefinite objects appear in the nominative. Although more nuanced definitions of specificity have been proposed (e.g., Fodor and Sag 1982; Enç 1991), we adopt the classical view proposed by Givón (1978), and associate specificity with identifiability of a referent by the speaker. With this in mind, consider that the NP object in (6) is specific, while the one in (7) is not. Regardless of this difference, both NPs, which are indefinite, must bear nominative case.

(6) Exu an jo-s (tu) kamnoni sa Adhana.
have.1SG a son-NOM (who) works in Adana.
'I have a son who works in Adana.' (specific)

(7) Irevu (an) toxtori-s, otis na in as en.
want.1SG a doctor-NOM whoever NA is HORT is
'I want a doctor, no matter who he is.' (non-specific)

In Turkish, DOM is triggered by *specificity* (Enç 1991; von Heusinger and Kornfilt 2005; Kornfilt 2009). While specific NP objects, regardless of their definiteness, receive accusative case in Turkish as in (8) and (9), non-specific NPs remain in bare form, as illustrated in (10) (Examples (8)–(10) are form von Heusinger and Kornfilt 2005, p. 5).[2]

(8) Ben kitab-ı oku-du-m.
I book-ACC read-PST-1SG
'I read the book.' (definite)

(9) Ben bir kitab-ı oku-du-m.
I a book-ACC read-PST-1SG
'I read a certain book.' (specific)

(10) Ben bir kitap oku-du-m.
I a book read-PST-1SG
'I read a book.' (non-specific)

To summarize, DOM in the two languages covers slightly different portions of the *Definiteness Hierarchy*. While in Turkish it is the specificity of an NP that triggers DOM, DOM in PhG appears only on definite NPs:

(11) PhG
Pronoun > Name > Definite > Indef. specific > Non-specific
Turkish

2. An anonymous reviewer points out to us the observation, made by Kornfilt (2009), that objects that carry an agreement morpheme receive accusative regardless of whether they are specific or not. We also observe similar facts for -mA nominalizations where neither agreement nor specificity is needed for accusative as in (i).

(i) yürü-me-yi sev-er-im
walk-NMLZ-ACC like-AOR-1SG
'I like walking.'

We acknowledge the extra complication and agree that a more precise generalization is needed. Our immediate observation is that accusative case correlates with the presence/absence of a D head/feature. The crucial point for us is that our proposed analysis can handle the facts regardless of what the precise generalization is. For expositional purposes, we make the simplifying assumption that accusative is correlated with specificity following Enç (1991).

That definiteness and specificity are closely related is not only implied by their linear adjacency on the *Definiteness Hierarchy* (11) but is also much discussed in the literature. Enç (1991), for instance, analyzes definiteness as an *identity link* relation between an NP and its discourse antecedent whereas *specificity* is identified as an *inclusion link* between the two. It is this close affinity between definiteness and specificity that has led previous authors (e.g., Karatsareas 2011, 2020) to claim that the DOM in PhG is due to some pattern replication from Turkish plus the association of the definiteness in PhG with specificity in Turkish. We take a slightly different view arguing that DOM in PhG emerged due to a substantial change in the input data received by the children acquiring the language. Before going into the details of this diachronic account, we would like to offer a formal account that unifies the case facts in PhG and Turkish.

3.2 DOM is structural accusative

In both Turkish and PhG, the DOM-case shows the typical properties of accusative case and it is structural. While the case in both languages has been traditionally labeled as *accusative*, we adopt the Dependent Case Theoretic definition of *accusative* to ensure that the DOM-case in both languages can be directly compared. Following Marantz (1991), we define accusative case as in (12).

(12) Accusative Case
Accusative is the case assigned to a lower NP in opposition to a higher NP in the same clause (V+I for Marantz (1991)).

DOM-case in both PhG and Turkish fits the definition in (12). Internal arguments of transitive clauses can bear accusative in both languages (13)–(14) (with the proviso that they carry the relevant feature on (11), to which we will turn in Section 4).

(13) Ece dergi-yi oku-du.
Ece journal-ACC read-PST
'Ece read the journal.' (Turkish)

(14) I liči skotsan ton apo-Ø.
the wolves kill.PST.3PL the.ACC fox-ACC
'The wolves killed the fox.' (PhG)

We follow Baker and Vinokurova (2010) and Baker (2015) in treating Dependent Case as structural. A common test for accusative being a structural case is case preservation under A-movement (Woolford 2006). Structural case is lost under A-movement. In both Turkish and PhG accusative is lost under passivization.[3]

(15) Dergi oku-n-du.
journal read-PASS-PST
'The journal was read.' (Turkish)

(16) Skotothin o apo-s s-is liči.
kill.PASS.PST.3SG the.NOM fox-NOM by-the wolves
'The fox was killed by the wolves.' (PhG)

4. Accounting for DOM

In this section, leaving DOM aside, we first present an account of *nominative, accusative*, and *dative* in both languages based on the *Case Realization Disjunctive Hierarchy* (Marantz 1991) (hereafter CRDH). Next, we modify the Dependent Case Theory to distinguish NPs based on their featural content to account for the DOM facts in the two languages. We call this 'Feature Relativized Dependent Case'.[4]

4.1 Nominative, accusative, and dative

One of the prominent views of case – *accusative* in particular – has been the CRDH, proposed by Marantz (1991, p. 24) (17). Marantz defines *accusative* as the dependent case assigned to a lower NP in opposition to a higher NP in the same clause. In this system, *nominative* is defined as the *unmarked* case within the clause. For the purposes of this paper, we only focus on the *dependent* case and the *unmarked* case.

(17) CRDH
a. Lexically governed case
b. Dependent case
c. Unmarked case (environment sensitive)
d. Default case

3. The original interpretation of the Case preservation under A-movement test presupposes that accusative was assigned and then lost. This is not something we adopt. In the Dependent Case view, accusative is not assigned at all. However, we think that the test is still valid in determining the structural status of accusative.

4. This notion came up in a conversation between Ümit Atlamaz and Mark Baker in Mark Baker's 2014 Seminar on Case. It has not been articulated elsewhere to the best of our knowledge.

The *dependent* case view of *accusative* case has recently seen a surge in adoption, especially through the work of Baker (2015), who extended it to account for *tripartite* case systems as well as structural *dative* as the dependent case assigned to the higher NP within the *v*P phase (see also Baker and Vinokurova 2010; Levin and Preminger 2015; Baker and Bobaljik 2017; Anagnostopoulou and Sevdali 2020 a.o.) In the following, we adopt Baker's implementation of the CRDH and propose the case rules in (18) and (19) to account for the distribution of *nominative, accusative* (and *dative* case) in Turkish and PhG.

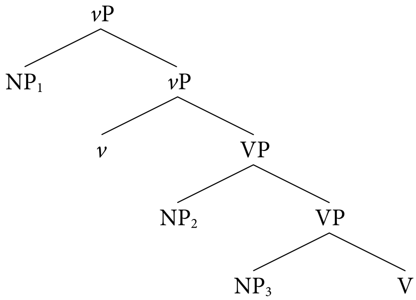
(18) Case Rules for PhG (first version)
  a. *Downward Dependent Case Rule*
     If NP1 c-commands NP2 in a clause, assign NP2 *accusative.*
  b. *Unmarked Case Rule*
     If an NP in a clause is not case marked otherwise, assign NP *nominative.*

(19) Case Rules for Turkish (first version)
  a. *Downward Dependent Case Rule*
     If NP1 c-commands NP2 in a clause, assign NP2 *accusative.*
  b. *Upward Dependent vP Case Rule*
     If NP1 c-commands NP2 in the *v*P, assign NP1 *dative.*
  c. *Unmarked Case Rule*
     If an NP in a clause is not case marked otherwise, assign NP *nominative.*

The case rules above account for the distribution of the *nominative, accusative*, and *dative* in Turkish and PhG, barring the DOM facts. Let us start with the Downward Dependent Case Rules ((18-a) and (19-a)) in both languages. Both rules predict that direct objects of transitive clauses should be *accusative* as already shown in (13) for Turkish and (14) for PhG. The rule also predicts that internal arguments will not receive accusative when they are not c-commanded by another NP in the clause. We have already shown that this is what we observe in passivization scenarios in (15) and (16).

A third prediction of the *Downward Dependent Case Rule* is that both of the internal arguments should receive accusative in a ditransitive clause. To illustrate this, consider the structure in (20), which provides the hierarchy among the arguments of a ditransitive clause. In this structure, the subject, NP_1, c-commands both the indirect object, NP_2, and the direct object, NP_3, and the indirect object, NP_2, c-commands the direct object, NP_3.

(20) Ditransitives



This is the point where Turkish and PhG diverge. While the prediction is borne out for PhG as illustrated in (21), in which both the direct and the indirect objects are marked in the accusative, double accusative is not possible in Turkish, as shown in (22-a). Instead, the indirect object must receive *dative* case as in (22-b).[5]

This follows from the presence of the *Upward Dependent vP Case* in Turkish which assigns *dative* to the higher NP that is inside the *v*P.[6]

(21) Dhoka paradha to nomati-Ø.
gave.1SG money the.ACC man-ACC
'I gave money to the man.' [PhG]

(22) a. *Ali Cem-**i** dergi-yi ver-di.
Ali Cem-ACC journal-ACC give-PAST
'Ali gave the journal to Cem.'

---

5. PhG and Turkish diverge in one further point: While indirect objects in Turkish receive *dative* case regardless of their feature content, in PhG definite indirect objects appear in the accusative, whereas indefinite ones appear in the nominative case and are further embedded under the preposition *s*- 'to'. Cf. (21) with (i):

(i) Dhoka paradha s-an nomati-s.
give.PST.1SG money to-a man-NOM
'I gave money to a man.' (PhG)

This means that DOM further extends to indirect objects in PhG as well. This is is precisely what our case rule in (23) predicts. In PhG, both of the internal arguments are subject to the downward dependent case rule leading to DOM in both direct and indirect objects. On the other hand, in Turkish, the upward dependent case rule is not feature relativized, yielding dative case on all goals regardless of their featural composition.

6. An anonymous reviewer asks what prevents the DO from receiving dative case. This simply follows from the definition of the dative rule as the upward dependent case in *v*P. Dative is only assigned to the higher of the two arguments.

b. Ali Cem-**e** dergi-yi ver-di.
Ali Cem-DAT journal-ACC give-PAST
'Ali gave the journal to Cem.' [Turkish]

To sum up the discussion so far, the facts that both *accusative* and *dative* in PhG and Turkish depend on the configuration of the NPs with respect to each other and with respect to a *v*P or the clause supports the Dependent Case account of these two cases in these languages.

Before closing this section, let us articulate our assumption regarding the *nominative* in PhG and Turkish. Following, Marantz (1991) and Baker (2015) we assume that the structural *nominative* is the unmarked case in the clause, regardless of its morphological nature. Any NP that has not received case otherwise will receive *nominative* as the unmarked case.

4.2 Feature relativized dependent case

The dependent case rules introduced in Section 4.1 account for the distribution of *accusative* and *dative* cases across clauses with different number of arguments (e.g., intransitive, transitive, causative, etc.). However, they fall short of accounting for the precise distribution of the *accusative* case which characterizes DOM in PhG and Turkish. To be more specific, the *Downward Dependent Case Rules* for PhG and Turkish are too coarse to account for the following facts.

1. Downward Dependent Case applies only to a subset of NPs.
   Pharasiot Greek :{pronoun, name, definite}
   Turkish :{pronoun, name, definite, specific}
2. Features that trigger DOM (accusative) are implicational.
   Pharasiot Greek : definite → name, pronoun
   Turkish : specific → definite, name, pronoun

A reference to features on the NP is thus required. Earlier accounts of DOM, such as Enç (1991), Næss (2004), and de Hoop and Malchukov (2008), provide identificational, hence feature-based analyses where a particular feature triggers a particular case (e.g. accusative). An often glossed over detail in feature-based analyses is that the they must make a covert assumption about which NP (subject or object) should receive DOM when it has the relevant feature (e.g. specific or definite). As a matter of fact, a particular feature does not guarantee an NP to receive a special DOM case unless it is also the object (or the subject if it is a Differential Subject Marking language).

To account for the precise distribution of accusative in PhG and Turkish, we relativize the *Downward Dependent Case Rules* for particular features. The following *Feature Relativized Downward Dependent Case Rules*, revised from (18-a) and (19-a) respectively, account for the distribution of DOM in PhG and Turkish.

(23) *Downward Dependent Case Rule for PhG* (final version)
If NP1 c-commands NP2 in a clause and NP2 contains [DEFINITE], assign NP2 *accusative.*

(24) *Downward Dependent Case Rule for Turkish* (final version)
If NP1 c-commands NP2 in a clause and NP2 contains [SPECIFIC], assign NP2 *accusative.*

With the *Downward Dependent Case Rules* modified as in (23) and (24), the case rules provided in (18)–(19) account for the precise distribution of *nominative, accusative*, and *dative* in both languages.

At this point one may raise the question whether feature relativization adds any further complexity to the Dependent Case Theory. The original version of the Dependent Case Theory, envisioned by Marantz (1991), is a context-sensitive grammar where context is the syntactic environment (e.g. clause). Extensions of the Dependent Case Theory to *dative* as in Baker (2015) build on this property of context-sensitivity and parameterize the context features for *accusative* in TP, *dative* in vP, etc. Our modification also builds on the context-sensitivity of the Dependent Case Theory and provides the feature content of an NP (a *case competitor*) as a context parameter to the Dependent Case Rules. In short, our modification does not change the essence of the Dependent Case Theory and utilizes mechanisms already available to the theory without any additional complexity.

The Feature Relativized Dependent Case mechanism we propose here seems to have an empirical advantage over the standard Dependent Case Theory, which was adopted to analyze DOM by Baker and Vinokurova (2010), and was further supported by Levin and Preminger (2015). Kalin and Weisser (2019) call this implementation the 'movement account of DOM' and we adopt the same term. Following Diesing (1992), the movement account builds on the assumption that *specific* NPs must move out of the VP to get interpreted. In a Phase-based syntax, movement is assumed to be to the *v*P phase edge where the specific object is in the same spell-out domain as the subject and triggers Dependent Case. On the other hand, a non-specific NP remains inside the VP resulting in the subject and the object remaining in different spell-out domains bleeding Dependent Case assignment. Relevant configurations are schematized in (25) and (26), where the grayed out part indicates the spell-out domain for the *v*P phase.

(25) [$_{TP}$ NP1 T [$_{vP}$ v [$_{VP}$ V NP2]]] (*non-specific NP2*)

(26) [$_{TP}$ NP1 T [$_{vP}$ NP2 v [$_{VP}$ V ~~NP2~~]]] (*specific NP2*)

The empirical motivation for the movement account comes from Sakha, where specific NPs precede VP-level adverbs (27-a) whereas nonspecific NPs cannot (27-b) (Baker and Vinokurova 2010). Relevant to this is the fact that specific NPs receive accusative but non-specific NPs remain bare (*nominative*).[7]

(27) *Sakha* (Baker and Vinokurova 2010, p. 602)
a. Masha salamaat-*(y) **türgennik** sie-te.
Masha porridge-ACC quickly eat-PST-3SG
'Masha ate the porridge quickly.'
b. Masha **türgennik** salamaat-(#y) sie-te.
Masha quickly porridge-ACC eat-PST-3SG
'Masha ate porridge quickly.'

Although the movement account is very attractive, it fails to capture two facts in our view. The first one comes from Turkish. Turkish is like Sakha in that specific NPs tend to precede adverbs like *hizlica* 'quickly' whereas non-specific NPs must follow them.

(28) a. Ali dergi-yi **hızlıca** oku-du.
Ali journal-ACC quickly read-PST
'Ali read the journal quickly.' (a specific journal)
b. Ali **hızlıca** dergi oku-du.
Ali quickly journal read-PST
'Ali read a journal quickly.' (a non-specific journal)

The crucial point is that specific NPs *can* precede adverbs like *hızlıca* but they do not have to: they can remain in situ and still be interpreted as specific as shown in (29). These state of affairs do not affect, to our mind, the interpretation of the adverb; i.e., in both the adverb quantifies over the process (cf. Cinque 1999, pp. 93, 103–104 and reference therein).

(29) Ali **hızlıca** dergi-yi oku-du.
Ali quickly journal-ACC read-PST
'Ali read the journal quickly.' (a specific journal)

7. Baker and Vinokurova (2010) report an additional complication by the possible presence of accusative on a non-specific NP that appears to remain in-situ as it follows the adverb *türgennik* 'quickly' in (27-b). They argue that presence of *accusative* on a non-specific NP triggers contrastive focus on the NP and implies a series of movement without explicating the details.

We take (29) to indicate that a specific NP can move out of VP but does not have to. Although this goes against Diesing's hypothesis, it is consistent with Holmberg's Generalization (Holmberg 1986). If specific NPs do not have to move out of VP, then the movement account of DOM fails to capture the accusative on the specific NP.

The second fact that challenges the movement account is that indefinite bare NPs in PhG do not bear accusative case. Under the assumption that specific NPs must move out of VP to get interpreted, we would expect PhG to behave like Sakha and Turkish and mark all specific NPs *including the indefinite ones* with *accusative*. This is not borne out, however. While indefinite specific NPs may appear to the left of VP-level adverbs, they are still marked in the nominative (V-to-T is assumed in the example below):

(30) Pitaksa a jo-s tarna sa Adhana.
send.PST.1SG a son-NOM quickly to.the Adana
'I sent one son (of mine) to Adana.'

We therefore conclude that *the Feature Relativized Dependent Case* mechanism can account for the precise DOM facts without facing the challenges faced by the movement account of DOM.

The final set of facts we want to address before turning to the next section is the implicational nature of DOM, as discussed in Section 3. Kalin (2018) derives the implicational hierarchy through a containment hypothesis. She proposes a range of heads (given in (31)), that are projected when a particular NP has the relevant features.

(31) *Heads in the Extended Nominal Projection* (Kalin 2018, p. 31)
a. Participant (semantically encoding first/second person)
b. Person (semantically encoding person)
c. Human (semantically encoding humanness)
d. Animate (semantically encoding animacy)
e. Name (semantically encoding the property of being a proper name)
f. Definite (semantically encoding definiteness)
g. Specific (semantically encoding specificity)
h. Number (semantically encoding number)

In this view, presence of certain features entails the presence of some other features. For example, the [PARTICIPANT] feature entails presence of [PERSON], [ANIMATE], and [HUMAN] features. Thus, if DOM is triggered by a feature (e.g. [SPECIFIC]) that is entailed by other features (e.g. [PARTICIPANT] or [DEFINITE]), then all the entailing features also trigger DOM by virtue of containing the triggering feature (e.g. [SPECIFIC]).

A similar account has been presented by Bárány (2017), who derives the implicational hierarchy from subset/superset relations. In this model, pronouns consist of the feature structures given in (32). Like in Kalin's proposal, if a language marks a 2nd person pronoun with {PARTICIPANT, π} features, then 1st person pronoun is also implied as it is a superset of the 2nd person.[8]

(32) [1] = {SPEAKER, PARTICIPANT, π }
[2] = {PARTICIPANT, π }
[3] = {π}

Following Kalin (2018) and Bárány (2017), we assume that the implicational hierarchy is a result of subset/superset relations among the feature compositions of NPs. This assumption, coupled with Feature Relativized Dependent Case mechanism, accounts for the distribution of accusative in PhG and Turkish. For ease of exposition and concreteness, we assume the feature composition in Table 3 for various NPs in Turkish and PhG.

Given Table 3, the feature relativized dependent case rules in (23) and (24) account for the distribution of accusative (i.e. DOM) in both languages.

Table 3. Feature composition of NPs

NP type	Featural composition
pronoun	{N, PRONOUN, DEFINITE, SPECIFIC}
definite NP	{N, DEFINITE, SPECIFIC}
indefinite specific NP	{N, SPECIFIC}
indefinite non-specific NP	{N}

To summarize this section, we have proposed a Feature Relativized Dependent Case mechanism to account for the distribution of *accusative* cases in PhG and Turkish to capture the DOM facts in both languages. The main question that we aim to answer in the rest of this paper is how contact with Turkish might have contributed to the emergence of DOM in PhG. In more precise terms, how did the downward dependent case rule in PhG change from the underspecified rule in (18-a) to the feature relativized rule in (23) without converging on the same DOM pattern with Turkish, i.e., (24)? While addressing the above question, we

8. Technically, the two proposals are slightly different as Kalin's proposal rests on containment of a particular feature (e.g.SPECIFIC), whereas Bárány's account allows combination of features (e.g. PARTICIPANT and π). In other words, in Bárány's account, it is possible for DOM to make reference to more than one feature rather than a single feature. However, the distinction is not crucial for the main proposal regarding containment and subset/superset relations.

make the following assumptions: the architecture of grammar is highly modular (Arregi and Nevins 2012), and each component of the grammar is dedicated to a particular task, with its own constraints. One of these components, the universal Case Module is guided by the CRDH (Marantz 1991). Since the DOM in both languages is associated with the accusative case, our particular focus is the Dependent Case Theory, which is what assigns accusative in the CRDH model. Convergence on the case system by two consecutive generations of speakers of the same language involves inducing the same set of dependent case rules and further associating these with the correct features based on the triggers in a finite amount of input produced by the previous generation. In other words, the Dependent Case Theory provides a set of possible rule types and therefore limits the space of possible hypotheses to be entertained by the learner. Heterogeneity in the input for the next generation, however it is introduced, may lead the learners in that generation to associate the relevant features with rules in a different way than the baseline grammar. In the next section, we will elaborate further on this view and offer our concrete analysis of the emergence of DOM in PhG.

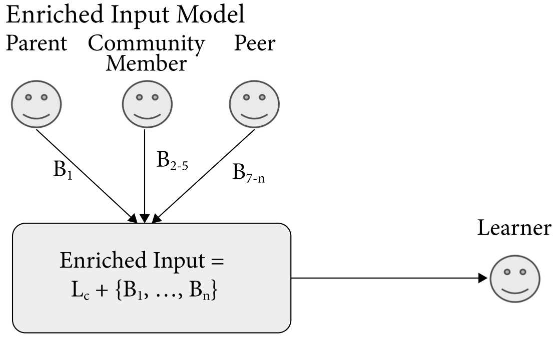
5. Language change as a learning problem

In line with much previous literature (Andersen 1973; Lightfoot 1979; Roberts 2021 a.o.), we hypothesize that language change begins at the level of an individual as a shift in frequencies with which a particular variant of linguistic expression is chosen over another. We further claim, again in line with previous literature, that such loss of equilibrium in an individual's grammar (I-language) emerges at the acquisition stage as an outcome of the opacity or ambiguity in the linguistic input (PLD) which this individual is exposed to. Faced with ambiguous input that does not trigger the same value as the one in the previous generation's grammar, or that is compatible with multiple hypotheses allowed by the general constraints of language design (including third factor), the individual may abduct a grammar that happens to be different than the one which underlies the input provided by the previous generation (generation here should be read as the system or systems underlying the input, PLD. PLD may obviously be produced by peers or older siblings as well as care-givers, such as parents). In short, language change begins because learners acquire a different grammar than the previous generation. This is the typical Z-model of language change, proposed by Andersen (1973) and adopted and modified later by Lightfoot (1979), Roberts (2021), and Kodner (2020) among many others. The Z-model is a reaction to the actuation problem, raised initially by Weinreich, Labov, and Herzog (1968, p. 102). One corollary of change from the perspective of acquisition is that change in I-language is 'catastrophic', 'saltational' (Lightfoot 1999) and 'abrupt' (Roberts 2021). *U-shaped learning* (Pinker and Prince 1988) shows that divergences arise in the acquisition process; they are common across many learners; and yet some may fail to result in language change. In this paper, we argue that, along with string ambiguity (resulting from loss of *morphophonological* cues, Roberts 2021), substantially limited input or enriched input also leads divergence (without ultimate convergence). These two types point to a well-established factor in change, namely language contact, which in turn assumes the existence of a multilingual context. Substantially limited input prototypically happens in the acquisition of heritage languages whereby the societal language grows more dominant than the ancestral/heritage one before the offset of the critical period. By enriched input we have in mind types of input that emerge under both recipient- and source-language agentivity, resulting in *borrowing* and *imposition* respectively (van Coetsem 1988, p. 25; Winford 2005, p. 377). The latter type appears in a wide range of contact scenarios, ranging from simple borrowing, to language shift, and to bilingual acquisition (see Muysken 2010 for different scenarios). Just like almost no single contact output can safely be explained with a unique contact mechanism, almost no contact scenario produces only one of the two types (which, for some authors, may characterize the same scenario, e.g., Seliger 1996). While we are aware of these shortcomings, given that some sort of bilingualism in PhG-Turkish is undeniable in the recent history of the speakers, we will hereafter used the term 'enriched input' in the context of PhG-Turkish contact.

The above framework answers how and, possibly, why a change emerges in the first place in a linguistic population but it does not address how this change, which initially emerges in a small number of speakers (perhaps one) escapes stochastic loss and propagates over a sizeable portion of the population so that it becomes a property of the E-language and is then traced by the linguists. Numerous studies since Kroch (1989) verify that propagation of a linguistic change is gradual and this gradualness concerns both individuals and aggregates of individuals (Pintzuk 1999, Harris and Campbell 1995, Lightfoot 2017, p. 525 a.o.). The gradualness may arise from both the nature of the grammar system, such as lexical diffusion etc, and, importantly, from 'use', i.e., the dynamics of contact between the innovators and those who come into contact with the innovators. Grammar competition model often attributes the success of the innovative form to some indexicality assumed by it (e.g., prestige) in the diglossic situation (Kroch 1989). Some alternative theories to language change such as *variational learning model* (Yang 2002) or *dynamical systems model* (Niyogi and Berwick 1997; Niyogi 2006) also address the issue of gradualness, offering intricate mathematical models; however, we think that such models can only be evaluated with large corpora that offer data from multiple stages of a language. Unfortunately, the written history of PhG does not extend beyond the mid 19th century, by which time DOM (and any other configuration that we deem relevant to the emergence of DOM) had been established. It is for this reason that we will have very little to say about how the innovative DOM pattern has become a property of the vernacular we know as PhG today.

In the rest of this paper, we model the emergence of DOM in PhG based on enriched input produced by a grammar that provides little cue for nominative-accusative distinction. We claim that there are two main ingredients in our analysis. First, the simplification/levelling in the nominal class system and loss of DP-internal agreement significantly reduced the frequency of triggers for successful acquisition of nominative-accusative morphology. Second, such input is also enriched with V-NP calques borrowed from Turkish that consist of a bare NP (no accusative) and some verb, such as *dhitu xarai* (lit. give face) < Turkish *yüz ver-* (lit. face give) 'pamper'. Initially, these calques were probably acquired as expressions without any production rule changes. These borrowed expressions become a part of the input data for a generation of children acquiring the language. When the number of these V-NP calques is below a certain threshold, they are kept as exceptions and memorized as lexical expressions (atomic or phrasal). Once the type frequency of these V-NP calques reaches a certain tipping point, the learner is forced to acquire a divergent set of production rules resulting in a divergent idiolect. When a substantial amount of learners in the community receive similar input, the language changes. Oversimplifying drastically, our proposal in a nutshell is this: Adults borrow fixed expressions, children turn them into a new rule. Our model is sketched in (33).

(33) Enriched Input Model



In (33), L_C stands for the language of the community and subscripted B represents a borrowed expression. In this model, input to the learner is considered as the totality of expressions provided by the parents, peers, and other members of the community. These expressions might be generated by distinct grammars, representing distinct idiolects. The borrowings may or may not have caused a rule change in the grammars of the borrowers. Even when an adult has borrowed a large amount of V-NP calques, we do not expect this to result in a production rule[9] if their grammar has reached maturation. However, this enriched input has the potential to lead to a divergent grammar for the acquirer, leading to language change along with the generation change.

In the following subsections, we first provide PhG-internal changes that are potentially relevant to the emergence of DOM. We later discuss the nature of V-NP calques in PhG, the essential culprits, and show how ubiquitous they are in the language. Then, we present the details of the learning algorithm we adopt. If we can show that an increased amount of V-NP calques, which contain bare NP objects instead of accusative objects, leads to a consistent emergence of DOM, our hypothesis is supported as a potential explanation for the emergence of DOM. This is what we do in the remaining sections: we present simulations where the learning algorithm is tasked to learn DOM in Turkish and PhG. We first show that the learning algorithm can consistently learn the Turkish DOM patterns by identifying the conditions under which the downward dependent case can apply. Next, we show that an increase in the type frequency of bare NP objects (V-NP calques) in a language like Greek results in DOM, as we hypothesized.

5.1 Triggers for DOM in PhG

PhG-type DOM is not attested in the earlier and attested stages of Greek, nor is it seen in any other extant mainland Greek dialects. In all the Greek varieties documented, accusative is the case appearing on direct objects without further constraints such as definiteness or specificity. In this section we elaborate on the argument that DOM is contact-induced by adding further details of this story. We use Modern Greek (MG) data to compare with PhG. We are aware of the fact that MG is neither an Asia Minor Greek dialect nor a predecessor of PhG; on the other hand, we believe that MG data represent, to a large extent, the relevant facts of PhG before the emergence of DOM, and in fact before contact with Turkish truly emerged (i.e., baseline PhG).[10]

---

9. A production rule is a generative mechanism that can generate expressions that go beyond the input.

10. Intensive contact between the Greek variety spoken in the region and Turkish must have been established somewhere between the 15th and the late 18th century. While there are undoubtedly many differences between MG and Post-classical/Medieval Greek (pre-15th c.), one could still take MG to be representative enough of those 'pre-contact' stages while presenting nominal inflection classes. The ideal scenario would of course be to provide information on the inflectional classes in post-Classical and Medieval Greek instead, which, however we cannot undertake here for reasons of space. We thank an anonymous reviewer for raising this issue.

5.1.1 *Cues for the accusative in MG…*

MG nominal stems are classified into inflectional classes (IC) based on presence or absence of stem allomorphy and the shape of the inflectional classes. Gender is commonly thought as part of the stem (see Ralli 2000 for details). MG morphologically distinguishes between nominative and accusative only on stems that belong to IC1 and IC2, i.e,. those that receive the inflectional endings *-os*, and *-is/-as/-us* respectively. IC1 hosts both masculine and feminine stems, whereas IC2 hosts only the masculine ones. In the singular, both ICs differentiate nominative from the accusative (Table 4).

Table 4. MG singular nouns (IC1–2)

	IC1a	IC1b	IC2
NOM	anthrop-os	am-os	maxit-is
ACC	anthrop-o	am-o	maxit-i
	'man' (masc)	'sand' (fem)	'warrior' (masc)

While IC1 further distinguishes between nominative and accusative in the plural, IC2 does not (Table 5).

Table 5. MG plural nouns (IC1–2)

	IC1a	IC1b	IC2
NOM	anthrop-i	am-i	maxit-es
ACC	anthrop-us	am-us	maxit-es
	'men' (masc)	'sands' (fem)	'warriors' (masc)

In all other inflectional class (IC3–8), which host feminine and neuter stems, nominative and accusative forms are syncretic in singular and plural numbers. In Table 6) we illustrate only the singular forms for space reasons.

Table 6. MG singular nouns (IC3–8)

	IC3	IC4	IC5	IC6	IC7	IC8
NOM	ghata	poli	avgho	spiti	kratos	soma
ACC	ghata	poli	avgho	spiti	kratos	soma
	'cat'	'city'	'egg'	'house'	'state'	'body'
	all feminine		all neuter			

Accusative and nominative cases are therefore not expressed overtly on singular and plural stems of IC3–4, and on plural items of IC2; however, the inflectional endings are not the only 'triggers' that signal Case. Morphological case appears on definite articles as well. In a definite DP, all stems of IC1–IC4 are accompanied by definite articles that make a formal distinction between nominative and accusative (Table 7). This means that case distinctions are syntactic but the stems in IC5–8 provide no morphological clue for this distinction.

Table 7. MG definite articles (IC1–4)

	Singular		**Plural**	
	MASC	**FEM**	**MASC**	**FEM**
NOM	o	i	i	i
ACC	ton	tin	tus	tis

In indefinite DPs, only masculine stems (i.e. those that belong to IC1 and IC2) are accompanied by an indefinite article that distinguishes between nominative and accusative, *viz. enas* (nom), *ena* (accusative). Whether the DP is indefinite or definite, given that adjectives agree with their head in gender, number and case, nominative vs. accusative case distinction can be signaled on at least some adjectival modifiers of items that belong to the inflectional classes IC1-IC4.

(34) a. o epimelis anthropos
the.M.NOM.SG diligent.M.NOM.SG man.M.NOM.SG
'the diligent man (nom)' (IC1)
b. ton epimeli anthropo
the.M.ACC.SG diligent.M.ACC.SG man.M.ACC.SG
'the diligent man (acc)' (IC1)

(35) a. i epimelis jineka
the.F.NOM.SG diligent.F.NOM.SG woman.F.NOM.SG
'the diligent man (nom)' (IC4)
b. tin epimeli jineka
the.F.ACC.SG diligent.F.ACC.SG woman.F.ACC.SG
'the diligent woman (acc)' (IC4)

When we turn to the members of IC5–8, which are all neuter, there is no case distinction signalled anywhere in the DP that host them. Neuter definite and indefinite articles (36-a),(36-b),(36-c), and neuter modifiers (36-d) do not encode accusative and nominative cases separately, similar to the neuter stems. The accusative and nominative forms of a (definite or indefinite) DP with a neuter noun are isomorphic (36-d) and (36-e).

(36) a. ena 'indefinite article; nom-acc'
b. to 'definite singular article; nom-acc'
c. ta 'definite plural article'; nom-acc'
d. to epimeles pedhi
the.N.NOM.SG diligent.N.NOM.SG child.N.NOM.SG
'the diligent child' (nominative)
e. to epimeles pedhi
the.N.ACC.SG diligent.N.ACC.SG child.N.ACC.SG
'the diligent child' (accusative)

Table 8 summarizes the appearance of nominative-accusative case distinction on different constituents of a DP that hosts any noun belonging to IC1–IC8. From a language acquirer's perspective, it is only when a (definite or indefinite) DP hosts a noun that is a member of IC5–8 that the same DP provides no clue (Lightfoot 1999; Lightfoot 2006) or (parameter) trigger (Clark and Roberts 1993; Roberts and Roussou 2003) to the acquirer for accusative-nominative case distinction.

Table 8. Nominative-accusative distinction visible in MG DP

		IC1	IC2	IC3	I4	IC5–8
on stems?	SG	yes	yes	no	no	no
	PL	yes	no	no	no	no
on def.art.?	SG	yes	yes	yes	yes	no
	PL	yes	yes	yes	yes	no
on indef. art.?	SG	yes	yes	no	no	no
on adj.?	SG	yes	yes	yes	yes	no
	PL	yes	yes	no	no	no

The above state of affairs summarized for MG do not exactly represent *every* fact of every stage in the history of Greek (for Ancient Greek, see e.g., Kakarikos (2009, 177ff), for post-Classical Greek, see Gignac (1981), form Medieval Greek, see Holton et al. (2019)); however, it does provide a good approximation of how agreement and case-expression is in every stage. With the above background, we now turn to see what changes have happened in PhG.

5.1.2 *…and their loss in PhG*

DOM is not the only innovation/change in PhG (see Bagriacik 2018 for many others), nor, we claim, is a phenomenon that emerged instantly. In particular, we maintain that there were three changes that have happened earlier or at least at the outset of the emergence of DOM in PhG.

Firstly, ICs were leveled in PhG in two significant ways: (a) no inflectional class includes stems of more than one gender (Bagriacik 2018). This means that either historical feminine stems in *-os* are retained as masculine in PhG; e.g., *o ammos* 'the sand (masc.)' (PhG) vs. *i ammos* 'the sand (fem.)' (MG), or such nouns were completely lost from the lexicon, e.g., *odhos* 'road (fem)' is no longer recognized by the speakers. (b) No stem of a specific gender combines in PhG with an inflectional ending that is also associated with a different gender. This refers to historical neuter nouns in *-os*, assigned to IC7 in MG. To the extent that these words are retained in PhG, they now appear as masculine and in IC1, e.g., the words *skevos* 'gadget', *kerdhos* 'gain', *xreos* 'debt', which are neuter (IC7) in MG, are masculine members of IC1 in PhG (as *škevos, čerdos*, and *xros* respectively). The ramification of (a) and (b) above is that the relationship between gender and ICs has become fully transparent. In this innovative IC system, the nominative-accusative distinction appears only on the *singular* members of IC1 and IC2. In Table 9 below, we provide the ICs corresponding to those of MG. IC4 and IC7 are not realized in PhG (cf. Table 6)

Table 9. PhG inflectional classes

		IC1	IC2	IC3	IC5	IC6	IC8
SG	NOM	apos	nomatis	neka	vo	pejkiri	goma
	ACC	apo	nomati	neka	vo	pejkiri	goma
PL	NOM	api	nomati	nečes	va	pejkire	gomata
	ACC	api	nomati	nečes	va	pejkire	gomata
		'fox'	'man'	'woman'	'egg'	'horse'	'piece'
		masculine		feminine	neuter		

Second, while the form of the definite articles remained identical to the inherited Greek ones (see Table 7), (36-b), (36-c)), the indefinite article is reduced to *a(n)* for all genders. This means that the indefinite article no longer encodes nominative and accusative distinction.

Finally, modifiers within (definite or indefinite) DP do no longer agree with the head noun in case and gender. They only encode number and appear in what is historically the neuter form. This means that adjectives (or any other modifiers) no longer encode case, even in DPs with masculine and feminine nouns.

The facts with respect to whether accusative-nominative is encoded at some point in a DP is summarized in Table 10.

Comparing Table 8 and Table 10, we reveal a crucial fact, which we think is one of the two main ingredients of the emergence of DOM; namely, the fact that

Table 10. Nominative-accusative distinction visible in PhG

		IC1	I2	IC3	IC5,6,8
on stems?	SG	yes	yes	no	no
	PL	no	no	no	no
on def. art.?	SG	yes	yes	yes	no
	PL	yes	yes	yes	no
on indef. art.?	SG	no	no	no	no
on adj.?	SG	no	no	no	no
	PL	no	no	no	no

accusative case is 'triggered' radically less in indefinite/bare contexts in PhG than in MG, and by assumption in the pre-contact version of PhG. This is so because neither the (a) indefinite article, nor (b) prenominal modifiers, nor feminine/neuter nouns (i.e., those that belong to IC3–8; Table 9) encode nominative vs. accusative case distinction. Added to this picture is the fact that, following various changes in the ICs and numerous nouns copied from Turkish, the ratio of neuter nouns, i.e., those that do not encode accusative-nominative distinction, to masculine (and feminine ones) increased dramatically. In a dictionary of PhG possibly from 1960s (Theodoridis undated), of the 3796 unambiguously nominal entries in the singular, 63% are in neuter gender, 26% in masculine and 11% in feminine. Furthermore, PhG underwent heteroclisis in its history; i.e., that mostly [-human] masculine singular nouns appear as neuter when in plural. These heteroclitic nouns make up nearly half of all the masculine entries (49%) in the same dictionary. If the dictionary is taken to reflect in some way the reality also prior to the emergence of DOM, we can safely conclude that 'accusative' is triggered much less in indefinite/bare contexts, and even more so if the noun in this context is in plural. From a learner's perspective, one could state that 71% indefinite/bare singular object DPs and 87% of indefinite/bare plural object DPs were weakly p-ambiguous, in the sense of Clark and Roberts (1993); i.e., such a string in the primary linguistic data constituted ambiguous trigger to the acquirer, expressing neither case value.

It is often claimed in non-variationist acquisition-based models of language change that parameter-ambiguity is the ingredient for the acquirer to 'reanalyze' the PLD and accordingly abduct a grammar that may deviate from the one that underlies the PLD (Lightfoot 1979; Clark and Roberts 1993; Roberts and Roussou 2003). Relevant to our case, Roberts (2021) further claims that the learning process in which the acquirer interrogates the PLD by looking for triggers, is optimized by a general, third-factor principle, namely:

(37) Input Generalization
If a functional head H_i of class C is assigned a formal feature *x*, assign *x* to all heads in C.

The learning path, according to Roberts, simply starts with no feature *x*. If *x* is detected on some head, then it is generalized to *all* heads. If in a later stage, a significant number of heads that do not bear *x* are detected, then a contrast among the maximal set of potentially *x*-bearing heads is made. This path is schematically shown in (38).

(38) a. default: $\neg\exists h[F(h)]$
b. $\exists h[F(h)] \rightarrow \forall h[F(h)]$
c. if $\exists$h $\neg[F(h)]$ is detected, select *h*' $\subset$ *h* as the domain of quantification and go back to (a). If no further *F*(*h*) is detected, stop.
d. If no further *F* (*h*) is detected, stop. (Biberauer et al. 2014, p. 111)

What would the reduced triggering of accusative vs. nominative distinction in indefinite DPs mean from this perspective? We think that once weak ambiguity arose with respect to nominative-accusative distinction in indefinite contexts, the acquirer faces the following possibility of reanalysis, which is in line with the above learning path and input generalization:

(39) the acquirer observes that nominative (subject) and accusative (object) forms of neuter and feminine DPs are non-distinct when they are indefinite. They generalize this non-distinctness to all indefinite DPs, including those with masculine nouns.

In other words, the acquirer may treat the non-distinctness as the default form. At this point, why the nominative form of the masculine ones is taken as the default form comes up as a question. According to Karatsareas (2011; 2020) the default choice of nominative is due to the fact that (adult, sentient) speakers of PhG identified the notion of indefiniteness in PhG with the notion of non-specificity of Turkish, the latter of which employs the default, nominative case. In an attempt to replicate the Turkish pattern, the proportional analogy dictated PhG speakers to choose the nominative as the default form to be employed in indefinite contexts in PhG as well.[11] While we think that Turkish influence is necessarily behind the

11. Spyropoulos (2020) offers an elegant account of differential argument marking in Asia Minor Greek. Relevant to our case is his analysis of Cappadocian DOM. According to Spyropoulos, accusative marking appears on definite NPs only post-syntactically, more specifically through an impoverishment rule, which involves the removal of [-oblique] from the case node relevant to the definite internal argument. In this schema, there is a retreat to a less specified exponence, namely to the zero exponent. In this study, we use 'default' simply referring to 'nominative', not in any technical sense.

choice of the nominative as the default, we differ from Karatsareas in the details, which we provide in the next section.

5.1.3 *Calquing from Turkish*

Recall from the previous section that PhG copied numerous nouns from Turkish (incidentally about 34% of all nouns in Theodoridis' dictionary are of Turkish origin). Nouns are not the only items that have been copied however. Being bilinguals in Turkish and Greek, such speakers introduced from Turkish almost all types of functional and lexical items (Bagriacik 2018). The important novelty for our case is the massive number of 'loan shifts/translations' (calques) as well as 'loanblends' (semi-calques) (Haugen 1950; Weinreich 1953, pp. 50–51) as V-NP fixed-expressions. There is a plethora of types of fixed expressions in languages that are composed of a lexical or light verb (V) and an object (NP) (Fraser 1970 and Everaert et al. 1995, pp. 3–4). These may be idioms, i.e., expressions that are completely semantically opaque and can only be interpreted locally, collocations, which allow meaningful syntactic operations while exhibiting some degree of non-compositionality (De Vries 2002, p. 78), or light verb constructions, which are semantically transparent and fully analyzable (Cattell 1981). V-NP fixed expressions that are claqued into PhG from their Turkish counterparts fall under all these categories. In most such cases, a native (i.e., Greek) verb combines with a noun of Turkish origin, which may not necessarily occur outside this configuration in PhG. The resulting V-NP chunk replicates the expression in Turkish, without however adopting the NP-V order (40):[12]

(40) a. ftenu <u>xaxi</u>
do right
'deserve' (PhG)

b. <u>hak</u> et-
right do
'deserve' (Turkish)

Such calques necessarily host a bare or an indefinite object noun in both languages. This means in PhG the direct object of such fixed expressions cannot be accompanied by a definite article and in Turkish they are necessarily in the default (nominative) from:

(41) a. ftenu (*to) xaxi
do the right

12. Considering that PhG is a VO language as opposed to Turkish, which is OV, this comes not as a surprise. In the case-theoretic perspective we adopt in this paper, word order differences plays no role.

b. hakk-(*i) et-
right-ACC do

The majority of calques (71%) in PhG host a direct object that is realized as a neuter or a feminine bare/indefinite noun as in (41-a), which, as discussed in the previous section, are not relevant to nominative/accusative case.[13] There are however another 29% of such calques with a bare/indefinite object masculine noun, which, as expected, appear in the nominative:

(42) a. ftenu jolči-s
make traveler-NOM
'send off sb' (PhG)
b. yolcu et-
traveler make
'send off sb' (Turkish)

It is quite possible that, while such V-NP fixed expressions were replicated in PhG by adult speakers, in some of them, the direct object may have been accommodated into the configuration in the nominative as a result of a (subconscious or conscious) attempt to replicate the form ensuring the same idiomatic meaning; yielding 'exact calquing' (Winford 2003, p. 98). Such V-NP expressions with the direct object in the nominative probably constituted p-strings in the same PLD, in which accusative is not triggered sufficiently on indefinite/bare objects (see Section 5.1.2). At earlier stages, the number of such V-NP chunks were probably not as high as they are listed in the dictionary today; however, as we show in Section 5.4, even a small number of such cases is enough to trigger nominative in *all* bare/indefinite direct object contexts across a few generations, in line with the Generalized Input factor.

We therefore reduce DOM in PhG to be an epiphenomenon on two (perhaps overlapping) developments in its history: (1) rise of p-ambiguity in terms of nominative-accusative in indefinite/bare direct object contexts, (2) creation of calques on the model of Turkish. On this point, we differ slightly from Karatsareas (2020): in our scenario, DOM was not borrowed; nor definiteness in PhG was matched with specificity in Turkish.[14]

13. V-NP fixed expressions counted here are taken from Theodoridis' manuscripts (Theodoridis 1939; 1960; 1964; 1966) and his dictionary (Theodoridis undated)

14. Spyropoulos' (2020) account of DOM in Cappadocian reaches the same conclusion as ours. In his account, however, /-s/ is not the nominative case form par excellence but rather an underspecified [-oblique] exponent, which would spell out case nodes that are traditionally associated with nominative and accusative. In this account, Cappadocian speakers do not use

We rather think that if some 'matching' occurred in (adult) bilingual minds at some point, it was due to the attempt of close approximation of calques in Turkish; which-at least in some cases-may have created V-NP chunks in which the bare/indefinite masculine direct object was realized in the nominative. Such V-NP chunks that deviate from the case facts of Greek further constituted triggers to the next-generation acquirers for generalizing the default nominative on all bare/indefinite masculine objects. This is in fact a conclusion, independently reached by Karatsareas (2020), for which he also provides circumstantial evidence: in another Asia Minor Greek dialect, namely the dialect of Silli (modern-day Konya), DOM is not recorded in any other configuration but in a few, isolated V-NP fixed expressions. The DOM facts of the dialect of Silli may therefore be taken to reflect a past stage of PhG.

5.2 Learning DOM in Turkish

In this section, we define the task for learning DOM in Turkish as acquiring a context sensitive dependent case rule relativized to a particular feature. Let us first list the Turkish patterns before defining the specific learning task we will be concentrating on.

We assume that the relevant features for case are {N, PRONOUN, DEFINITE, SPECIFIC} and the simplest nominal expression with no additional discourse features has the category feature N. Specific nominals have an additional SPEC feature, and other features keep stacking as in Table 11. A learner interacting with Turkish will receive random input data drawn from the patterns described in Table 11. The task is to acquire a grammar that generalizes from the random input and produces the same patterns on unseen data. The ultimate task is to converge on a dependent case rule like the one in (43).

Table 11. Case patterns in Turkish

NP type	Featural composition	Case
pronoun	{N, PRONOUN, DEFINITE, SPECIFIC}	ACCUSATIVE
definite NP	{N, DEFINITE, SPECIFIC}	ACCUSATIVE
indef. spec. NP	{N, SPECIFIC}	ACCUSATIVE
indef. non-spec. NP	{N}	NOMINATIVE

bare/indefinite objects in the nominative but in rather an underspecified case form. See Spyropoulos (2020, Section 4) for details.

(43) *Downward Dependent Case Rule Turkish*
If NP1 c-commands NP2 in a clause and NP2 contains [SPECIFIC], assign NP2 *accusative.*

Learning a rule like (43) involves learning two components, which can be modeled together or independently. These are: (1) learning that Turkish has a downward dependent case rule (2) learning that the dependent case rule applies only when the object is specific.[15] While the first component determines the presence of *accusative* in the language, the latter component regulates its distribution based on the featural content of the object, i.e. DOM.

As our focus is on DOM, we will make the simplifying assumption that learning DOM is learning the conditions under which the downward dependent case applies. Thus, the learning task is defined as in (44).

(44) *Learning Task for DOM in Turkish*
Find the contextual features where the Downward Dependent Case Rule applies.

Now that the learning task is defined, we discuss the learning algorithm we adopt and then provide a description of the simulated input data. Then we apply the model on the data and discuss the results.

5.2.1 *The learning algorithm*

The learning task defined above is a classification task where the input is a set of noun phrases with an array of abstract features and the output is a decision regarding whether the downward dependent case rule applies or not. The goal for the learner is to find the set of features that designate the context for the downward dependent case to apply *productively*. While in principle the classification task can be modeled with a range of machine learning algorithms (e.g. naive-bayes classifier, logistic regression, support vector machines, etc.), we adopt the Greedy and Recursive Search for Morphological Productivity (henceforth ATP)[16] algorithm by Belth et al. (2021).

ATP is a decision tree model based on an abductive reasoning procedure. It takes input a set X of instances in the form (*ℓ, F*, ı), where *ℓ* is a *lemma, F* ⊆ Ω is a set of features from the feature space Ω, and ı is an inflected form corresponding

15. There are other contextual restrictions such as being in a clause as opposed to some other syntactic environment but we gloss over those as they are not directly relevant to DOM phenomena.

16. Belth et al. (2021) call the main search procedure in the model as the *Abduction of Tolerable Productivity* as it involves Peircean abductive learning in combination with the Tolerance Principle (Yang 2016).

to lemma ℓ and features F. An example we quote from Belth et al. (2021) is given in in (45).

(45) (*walk*, {3, SG, PAST}, *walked*) (Belth et al. 2021, p. 2)

In (45), ℓ = *walk*, F = {3, SG, PAST}, and ι = *walked*. This is a supervised model where the input is a lemma and a set of abstract features and the output is an inflected surface form. Our implementation diverges from the implementation by Belth et al. (2021) slightly in that our output is not an inflected form but a signal for the dependent case module. In other words, our model does not return inflected surface forms but a decision regarding whether the dependent case should apply or not. Our instances look like the representations in (46).

(46) a. (NP_i, {N, DEF, SPEC}, NP_i – **ACC**)
b. (NP_j, {N}, NP_j)

The NP_n in the first index of the representation is a noun phrase type (not a token) where the subscript n indicates the index of the type. For example, NP_1 could be the noun phrase *iki kedi* 'two cats' while NP_2 could be *çocuklar* 'kids'. Thus, in our implementation ℓ is not a morphological *lemma* but the string representation of a syntactic phrase. The feature sets are abstract syntactico-semantic features $F \subseteq \Omega$ associated with the NP_n and the feature space (Ω) for Turkish is given in (47).

(47) Feature Space for Turkish
Ω = {N, PRN, DEF, SPEC}

Finally, the third indices of the triplet representations in (46) are the output forms that signal whether the dependent case can be applied or not. Since in both Turkish and Greek, the dependent case is downward, i.e. *accusative*, we use ' -ACC' as the diacritic indicating that dependent case applies for a given instance. The absence of the '-ACC' diacritic signals that the dependent case cannot apply. So, in (46-a), the NP_i is a pronoun carrying the features {N, PRN, DEF, SPEC} and the dependent case applies when this NP is in the object position. On the other hand, in (46-b), NP_j is an indefinite non-specific noun phrase with the feature {N} and the dependent case does not apply for this NP.

ATP recursively builds a decision tree. At each recursive step, ATP starts with the most common affix (a diacritic in our implementation). Among the features that are associated with the instances that has this affix (diacritic), it searches for the feature that will maximize consistency based on the formula in (48).

(48) *Feature to split on* (Belth et al. 2021, p. 2)

$$\hat{s} = arg\ max_{s \in \Omega} = \frac{f_{\sigma max}(\chi_s)}{|\chi_s|},$$

where $\chi_s = \{(\ell, F, \imath) \in \chi : s \in F\}$ is the set of instances with feature $_s$
and $f_{\sigma\max}.(\chi_s)$ is the frequency of the most frequent suffix in χ_s

Once a feature is selected, the instances are split into two consisting of χ_s (instances that have the feature *s*) and $\neg\chi_s$ (the complement set of χ_s), resulting in a binary branching tree with two leaves. The path from the root of the decision tree to a leaf is interpreted as a *rule*. Each rule is checked for *productivity* where productivity is determined by the Tolerance Principle (Yang 2016) defined as in (49).

(49) *Tolerance Principle* (Yang 2016, p. 2)
r is productive iff

$$e \leq \theta_N \triangleq \frac{N}{lnN},$$

where *r* is a rule
e is the number of exceptions
and *N* is the number of instances in the scope of the rule

When the number of exceptions for a rule is smaller than N divided by the natural log of N, the rule is maintained as productive, otherwise, it is discarded. ATP applies recursively (creating new splits when no productive rule is found and increasing the depth of the tree) until a maximally productive rule is found and till all the features are traversed. When no productive rule is found by the time the entire feature space is traversed, the search for productivity fails and the training instances are memorized without any generalization. See Belth et al. (2021) for further details on ATP.

ATP is designed as a deterministic algorithm where the decision regarding which feature to split on at each recursive step is determined by the algorithm in (48) and the test for productivity is also deterministic due to the nature of Tolerance Principle given in (49). Nevertheless, there are certain edge cases causing some non-determinacy resulting in multiple models accounting for the same data. One such edge case is when a single affix (diacritic) can be associated with two distinct features where either of the features can maximize consistency (e.g. $arg\ max_{s \in \Omega} = 1$ for both features). Under such circumstances, ATP will pick a feature randomly and this randomness will yield two distinct models that take the same input and return the same output.[17]

We have adopted ATP for both theoretical and practical reasons. On the theoretical side, Tolerance Principle provides a quantifiable model of productivity as a tipping point where observations are turned into productive rules once a certain amount of input is received and the number of exceptions are below a certain threshold (Yang 2016). Belth et al. (2021) have demonstrated that the learning trajectories of the ATP model resemble the developmental paths of children displaying similar errors and generalizing over wug tests.[18] On the practical side, Belth et al. (2021) provide a Python implementation of the ATP, which made it possible for us to run simulations by providing relevant input data and some minor modification on the code.[19]

5.3 Turkish simulations

The goal of Turkish simulations is to test the feasibility of learning DOM, i.e. the contextual features triggering the downward dependent case, using the ATP algorithm. Since we do not have access to a child directed speech (CDS) corpus, we used simulated data with varying proportions of definite, specific, and indefinite-non-specific NP type frequencies to identify the kind of input needed to learn DOM in Turkish. We ran two types of simulations: (1) 100-NP simulations and (2) realistic data simulations.

5.3.1 *100-NP simulations*

Our first set of simulations aimed to test the proportions of definite, indefinite-specific, and non-specific NP types required to learn DOM when the type frequency is 100. In other words, when a child receives 100 NP types (not tokens) as input, what proportions result in learning DOM with a 100% accuracy. Each simulation dataset consists of 106 type instances with 6 of them being types representing pronouns and some permutation of definite, indefinite-specific, and non-specific nominals. We tested all 4581 permutations where each variable has at least 1 instance. Table 12 illustrates the dataset permutations.

---

17. We thank Caleb Belth for helping us figure out the source of non-determinacy in the Pythonic implementation of ATP (https://github.com/cbelth/ATP-morphology) by Belth et al. (2021).

18. See Belth et al. (2021) for further discussion.

19. We only switched off the phonological feature extraction module of the code to avoid phonological features to be considered when the model is being built. Code for our simulation experiments can be found at https://github.com/umitatlamaz/dom-simulations.

Table 12. Data permutations

Dataset #	Pronoun	Definite	Specific	Bare	# of training iters.
1	6	1	1	98	100
2	6	1	2	97	100
3	6	1	3	96	100
…					100
4579	6	96	1	3	100
4580	6	97	1	2	100
4581	6	98	1	1	100

The ATP model was trained for 100 iterations on each dataset, where each iteration represents an independent learner. The main goal was to simulate 100 children receiving the same input and measuring the proportion of learners who acquire a grammar that can achieve 100% accuracy on unseen data.[20, 21] In each iteration, we tested the accuracy of the model by comparing its predictions on unseen test data. The test data consists of four logically possible patterns, described in Table 11. A model is deemed successful if it acquires a grammar that predicts the correct distribution of accusative case on all four patterns (i.e. 100% accuracy). Otherwise, it is considered a failure. A dataset (with a particular permutation) is considered *representative* if each of the 100 learners (i.e. iterations) achieves 100% accuracy on the test set. If some (but not all) of the learners achieve 100% accuracy, the dataset is considered to be *ambiguous*. Otherwise, the dataset is considered to be *not representative* of Turkish. Table 13 summarizes the results.

Table 13. Turkish 100-NP simulation results summary

Dataset status	Num. of datasets	% of datasets
representative	1863	38.4%
ambiguous	943	19.4%
not representative	2045	42.2%
total	**4581**	**100%**

20. Note that the model has some degree of non-determinacy stemming from the random feature selection when the decision tree is being built.

21. As an anonymous reviewer points out, this is an idealization as no two children are guaranteed to receive exactly the same input.

When the type frequency of the input is kept constant at 100, 38.4% of all the possible permutations leads to successful convergence on Turkish DOM whereas 19.4% lead to a partial convergence where some learners acquire the Turkish DOM patterns while the others fail to do so. When the data is ambiguous, on average 49% of the learners reach 100% accuracy on the test data, while 51% of the learners fail to converge on Turkish, leading to a divergence in a group of learners. In the remaining 42.2% of the permutations, there is no convergence on Turkish DOM patterns.

Keeping the type frequency constant at 100, only 38.4% of the data guarantees convergence on Turkish DOM patterns. However most (probably all) of the children acquiring Turkish converge on the Turkish DOM patterns. Thus, the input they receive must be within a certain distribution. The green area on the ternary plot in Figure 1 describes the input zone that is *representative* of Turkish, the blue area indicates the *ambiguity* zone, and the red areas indicate the *not representative* zone.[22]

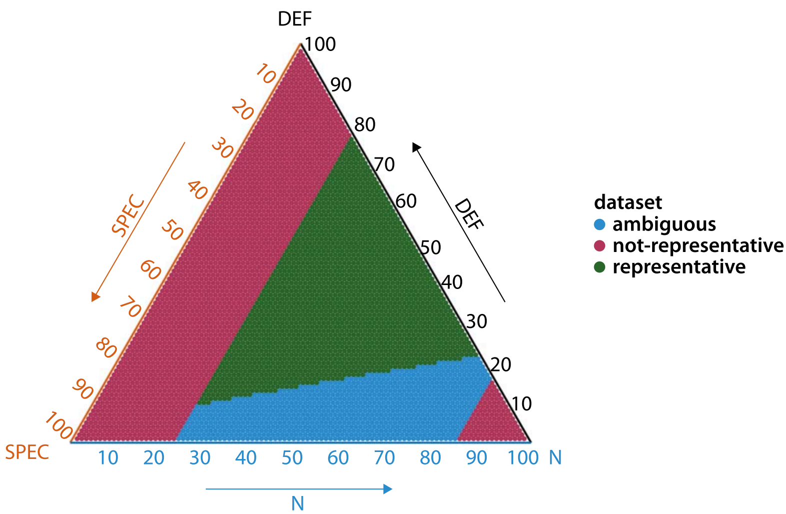


Figure 1. Representativeness of permutations

As long as the distribution of the features are not at extremes (e.g. only few instances or too many instances), the model is able to learn DOM in Turkish. This also aligns with the fact that most (random) corpora follow a certain distribution (e.g. Zipfian) but not the extreme distributions. Unfortunately, we do not have any CDS corpus where we could compare the distributions depicted in Figure 1 for

22. Our use of ternary plots is inspired by Kodner (2020).

the first 100 NPs and test whether they overlap with realistic input. However, in the next section, we report the results of a second type of simulations where we used some distribution from a realistic corpus.

5.3.2 *Realistic data simulations*

In an ideal scenario, we would model Turkish DOM using corpora consisting of child directed speech (CDS). Nevertheless, we do not have such resources. The frequency of lexical items in corpora have been shown to have a Zipfian distribution. Yang (2018, pp. 4–6) and Kodner (2020) have shown that trimming the low frequency words from the adult literary genres brings the corpus close to CDS vocabulary to a degree that the difference is insignificant in most cases.

Following this insight, we used a literary corpus consisting of 3070 books from a wide range of genres to get an approximate distribution of accusative and bare NP objects. We parsed the corpus using the Stanza library (Qi et al. 2020) to extract the NP objects.[23] We successfully parsed about 260K sentences with an NP object with the following distributions.

Table 14. Literary corpus NP distributions

NP type	**%**
definite accusative	34%
indefinite accusative	9%
no accusative	57%

Using the distributions in Table 14, we ran 99 simulations. In each simulation, we incremented the total type frequency by 100 and ran 100 iterations (learners). Our results showed that when the type frequency is 100, the data is ambiguous leading to a 41% convergence rate (i.e. 41 learners acquired DOM while the others failed). However, once the total type frequency reaches 2400, convergence reaches 100%. Our simulation results are plotted in Figure 2.

Our results show that ATP can learn DOM in Turkish once the total type frequency reaches a sizeable number (2400 in this case) even though the relative proportion of indefinite specific NPs is quite low compared to definite and non-specific NPs. Now that we have shown that ATP is capable of learning the distribution of accusative case (DOM), we move on to show how an increase in the type frequencies of bare NPs with a slightly different feature space (Ω) leads to the PhG DOM.

23. We thank Karahan Şahin for sharing the outputs of the parser with us.

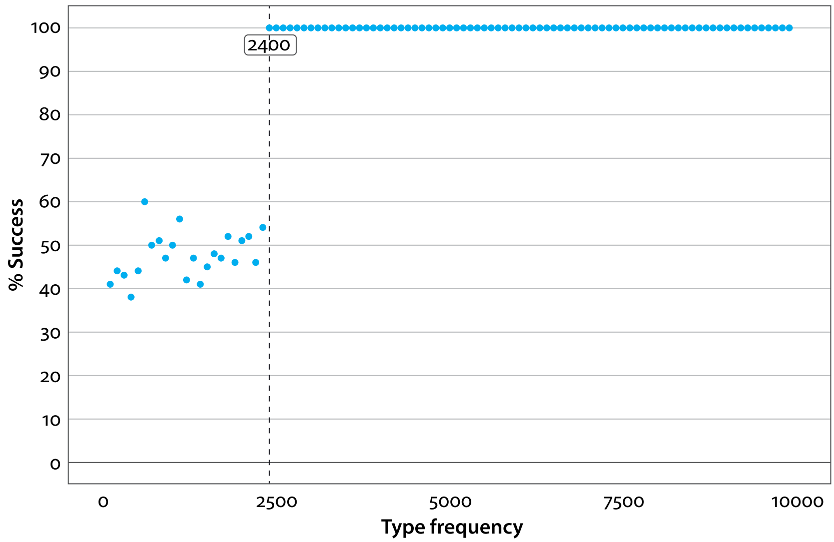


Figure 2. Turkish type frequencies

5.4 PhG simulations

In this section, we discuss the learning task for the PhG speakers and the results of the 100-NP simulations for PhG. We show that once the type frequency of the bare NP objects reaches a certain threshold, DOM emerges consistently. Let us first define the learning task and the input space. Unlike in Turkish, specificity does not play a role in the distribution of accusative in PhG. Thus, we have defined the feature space as in (50).[24]

(50) $\Omega = \{$N, PRONOUN, DEFINITE$\}$

Given the feature space in (50), the typology of data a PhG speaking child receives is given in Table 15.

The child's task is to learn the correct set of contextual features for the Downward Dependent Case (accusative). Oversimplifying, in baseline Greek, objects are consistently accusative marked. Some exceptions are tolerable and they are simply memorized without a consistent DOM grammar. Thus our assumption is that the type frequency of the last row (nominative objects) was at or under a *tol-*

24. We have also experimented with a feature space including the feature specific. The results were unsuccessful in that we never reached 100% learner convergence. Since specificity is not a good predictor of accusative case in PhG, we removed it from the feature space. This is consistent with the theoretical proposal that it is definiteness but not specificity that matters for DOM in PhG.

Table 15. Case patterns in PhG

NP type	Featural composition	Case
pronoun	{N, PRONOUN, DEFINITE}	ACCUSATIVE
definite NP	{N, DEFINITE}	ACCUSATIVE
indefinite NP	{N}	ACCUSATIVE
indefinite NP	{N}	NOMINATIVE
		⇑
		calques & case neutralization

erable threshold in the baseline speakers before the emergence of the DOM. V-NP calques increased the type frequency of nominative objects resulting in either ambiguous input or an unambiguously different input causing the emergence and propagation of DOM.

We used the same 100-NP simulation settings as in Turkish. A total of 4581 datasets with all the permutations of the features and their type frequencies were used. For each dataset, we trained 100 independent learners and tested their predictions on a novel test set consisting of definite NPs and indefinite NPs. The results are summarized in Table 16.

Table 16. PhG 100-NP simulation results summary

DOM status	Num. of datasets	% of datasets
no emergence	3851	79.3%
emergence	90	1.9%
convergence	910	18.8%
total	**4581**	**100%**

As presented in Table 16, in 79.3% of the datasets, PhG type DOM does not emerge. 1.9% of the datasets are *ambiguous* resulting in emergence of DOM in some learners but never reaching a complete convergence among the learners with an average rate of 51.2%. Finally, 18.8% of the datasets are unambiguous and result in complete convergence among the learners resulting in 100% of the learners acquiring a grammar producing the PhG patterns. The results are plotted on the ternary plot in Figure 3. Figure 3 shows the DOM zone, which indicates that a higher number of bare NP objects leads to the acquisition of a grammar with DOM. The green area indicates the datasets leading to a convergent acquisition of DOM, the red area indicates the ambiguous zone where some of the learners acquire DOM, and the blue area indicates the permutations where DOM does not emerge.

To make the evaluation results a little more interpretable, we have also ran *evtree*, an evolutionary CART algorithm (Grubinger, Zeileis, and Pfeiffer 2014) to identify the convergent DOM zones in the 100-NP Greek simulations. The results are plotted in Figure 4.

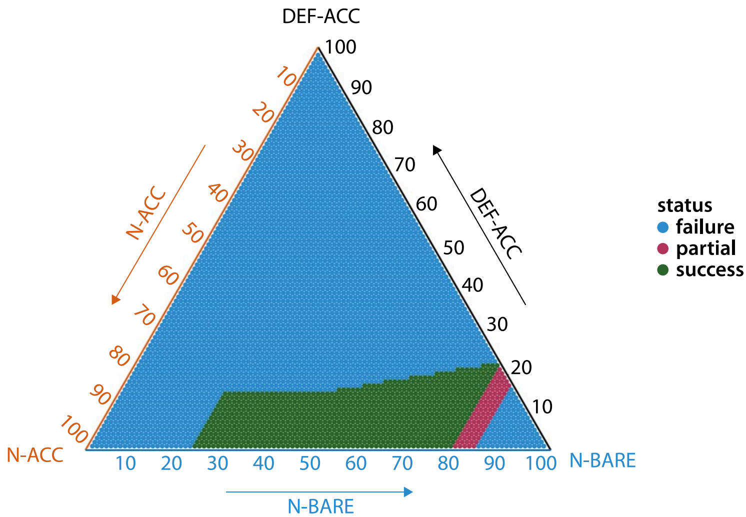


Figure 3. PhG simulation results

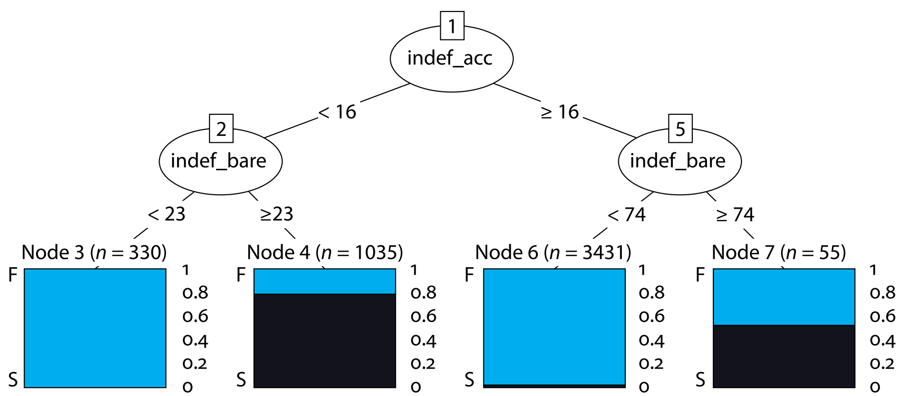


Figure 4. Evolutionary tree analysis of Greek simulations

The *evtree* model reveals that the majority of the convergent DOM simulations occur when the number of bare indefinites is at or above 23 and the number of accusative indefinites is below 16.

6. Conclusion

In this paper, we discussed DOM in PhG as an emergent phenomenon due to contact with Turkish. We first presented an in depth comparison of DOM in Turkish and PhG, concluding that they are both instances of structural accusative case and analyzed them as context sensitive dependent case. We hypothesized that the emergence of DOM in PhG is due to an increased number of V-NP calques borrowed from Turkish. These V-NP calques consist of a bare NP object without any indicator of accusative case. Using a generalization learner, ATP (Belth et al. 2021), we simulated the learning process of accusative case in Turkish and PhG. Our results support the hypothesis that an increased number of bare NP objects results in PhG-type DOM.

Acknowledgements

We would like to thank Deniz Tat, Muhammed İleri, Karahan Şahin, Caleb Belth, Charles Yang, Jordan Kodner, Pavel Logačev, and the audience at the Workshop 'Formal Approaches to Contact in/with Turkish' held at UiT The Arctic University of Norway (Tromsø) on May 19–22, 2022 for their insightful comments and support at various stages of this work. Special thanks to two anonymous Linguistic Variation reviewers whose feedback improved the paper significantly. We acknowledge the Boğaziçi Research Fund for the financial support to our respective projects (Ü. Atlamaz: 19348; M. Bagriacik: 17921). All errors are our own.

References

doi Aissen, Judith (2003). "Differential object marking: Iconicity vs. economy". In: *Natural Language & Linguistic Theory* 21.3, pp. 435–483.

doi Anagnostopoulou, Elena and Christina Sevdali (2020). "Two modes of dative and genitive case assignment: Evidence from two stages of Greek". In: *Natural Language & Linguistic Theory* 38.4, pp. 987–1051.

Anastasiadis, Vasilis (1976). "I sintaksi sto Pharasiotiko idioma tis Kappadokias [The syntax of the dialect of Pharasa in Cappadocia]". PhD thesis. University of Ioannina.

doi Andersen, Henning (1973). "Abductive and deductive change". In: *Language*, pp. 765–793.

Andriotis, Nikolaos (1948). *To glossiko idioma ton Pharason* [*The dialect of Pharasa*]. Athens: Ikaros.

doi Arregi, Karlos and Andrews Nevins (2012). *Morphotactics: Basque auxiliaries and the structure of spellout*. Dordrecht: Springer.

Bagriacik, Metin (2018). "Pharasiot Greek: Word order and clause structure". PhD thesis. Ghent University.

doi Baker, Mark C. (2015). *Case: Its principles and its parameters*. Cambridge, UK: Cambridge University Press.

doi Baker, Mark C. and Jonathan David Bobaljik (2017). "On inherent and dependent theories of ergative case". In: *The Oxford handbook of ergativity*. Ed. by Jessica Coon, Diane Massam, and Lisa deMena Travis. Oxford: Oxford University Press, pp. 111–134.

Baker, Mark C. and Nadezhda Vinokurova (2010). "On tense and copular verbs in nonverbal predications in Sakha". In: *Rutgers Working Papers in Linguistics 3*. Ed. by Peter Staroverov. New Brunswick, NJ: Rutgers Department of Linguistics, pp. 31–63.

Bárány, András (2017). *Person, case, and agreement: The morphosyntax of inverse agreement and global case splits*. Oxford: Oxford University Press.

Belth, Caleb et al. (2021). *The Greedy and recursive search for morphological productivity*. arXiv: 2105.05790v1 [cs.CL].

doi Biberauer, Theresa et al. (2014). "Complexity in comparative syntax: the view from Modern parametric theory". In: *Measuring grammatical complexity*. Ed. by Frederick J. Newmeyer and Laurel B. Preston. Oxford: Oxford University Press, pp. 103–127.

Cattell, Ray (1981). *Composite predicates in English*. Sydney: Academic Press.

doi Cinque, Guglielmo (1999). *Adverbs and functional heads. A Cross-linguistic Perspective*. Oxford: Oxford University Press.

Clark, Robin and Ian Roberts (1993). "A computational approach to language learnability and language change". In: *Linguistic Inquiry 24*, pp. 299–345. url: https://www.jstor.org/stable/4178813

Comrie, Bernard (1979). "Definite and animate direct objects: A natural class". In: *Linguistica Silesiana 3*, pp. 13–21.

Croft, William (1988). "Agreement vs. case marking and direct objects". In: *Agreement in natural language: Approaches, theories, descriptions*. Ed. by Michael Barlow and Charles A. Ferguson. Stanford, CA: CSLI, pp. 159–179.

Dawkins, Richard M. (1916). *Modern Greek in Asia Minor: A study of the dialects of Silli, Cappadocia and Pharasa, with grammar, texts, translations and glossary*. Cambridge, UK: Cambridge University Press.

doi Dawkins, Richard M. (1940). "The dialects of Modern Greek". In: *Transactions of the Philological Society 39*, pp. 1–38.

doi de Hoop, Helen and Andrej L. Malchukov (2008). "Case-marking strategies". In: *Linguistic Inquiry 39*, pp. 565–587.

Diesing, Molly (1992). *Indefinites*. Cambridge, MA: The MIT Press.

Enç, Mürvet (1991). "The semantics of specificity". In: *Linguistic inquiry 22*, pp. 1–25. url: https://www.jstor.org/stable/4178706

Everaert, Martin et al., eds. (1995). *Idioms: Structural and psychological perspectives*. Hillsdale, NJ: Erlbaim.

Favis, Vasilis (1948). "Sintaktike paratirisis is to idioma ton Farason [Syntactic observations on the dialect of Pharasa]". In: *Epetiris Vizantinon Spoudon 18*, pp. 173–191.

doi Fodor, Janet D. and Ivan A. Sag (1982). "Referential and quantificational indefinites". In: *Linguistics and Philosophy 5*, pp. 355–398.

Fraser, Bruce (1970). "Idioms within a Transformational Grammar". In: *Foundations of Language 6*, pp. 22–42. url: https://www.jstor.org/stable/25000426

Gignac, Francis (1981). *A Grammar of the Greek papyri of the Roman and Byzantine periods*. Vol. 2: Morphology. Milano: Cisalpino-La Go-liardica.

Givón, Talmy (1978). “Definiteness and referentiality”. In: *Universals of human language*. Ed. by Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravcsik. Vol. 4. Stanford, CA: Stanford University Press, pp. 291–330.

doi Grubinger, Thomas, Achim Zeileis, and Karl-Peter Pfeiffer (2014). “evtree: Evolutionary learning of globally optimal classification and regression trees in R”. In: *Journal of Statistical Software 61*, pp. 1–29.

doi Harris, Alice and Lyle Campbell (1995). *Historical syntax in cross-linguistic perspective*. Cambridge, UK: Cambridge University Press.

doi Haugen, Einar (1950). “The analysis of linguistic borrowing”. In: *Language 26*, pp. 210–213.

Holmberg, Anders (1986). “Word order and syntactic features in the Scandinavian languages and English”. PhD thesis. Stockholm University.

doi Holton, David et al. (2019). *The Cambridge grammar of Medieval and Early Modern Greek*. Vol. 2, 4. Cambrdige, UK: Cambridge University Press.

doi Janse, Mark (2004). “Animacy, definiteness, and case in Cappadocian and other Asia Minor Greek dialects”. In: *Journal of Greek linguistics 5*, pp. 3–26.

Kakarikos, Konstantinos (2009). “Morphological and “semantic” examination of the case system of Ancient Greek”. PhD thesis. University of Athens.

doi Kalin, Laura (2018). “Licensing and Differential Object Marking: the view from Neo-Aramaic”. In: *Syntax 21*, pp. 112–159.

doi Kalin, Laura and Philipp Weisser (2019). “Asymmetric DOM in coordination: A problem for movement-based approaches”. In: *Linguistic Inquiry 50*, pp. 662–676.

Karatsareas, Petros (2011). “A study of Cappadocian Greek nominal morphology from a diachronic and dialectological perspective”. PhD thesis. University of Cambridge.

doi Karatsareas, Petros (2020). “The Development, preservation and loss of Differential Case Marking in Inner Asia Minor Greek”. In: *Journal of Language Contact 13*, pp. 177–226.

Kodner, Jordan (2020). “Language acquisition in the past”. PhD thesis. University of Pennsylvania.

doi Kornfilt, Jaklin (2009). “DOM and two types of DSM in Turkish”. In: *Differential Subject Marking*. Ed. by Helen de Hoop and Peter de Swart. Dordrecht: Springer, pp. 79–111.

doi Kroch, Anthony (1989). “Reflexes of grammar in patterns of language change”. In: *Language Variation and Change 1*, pp. 199–244.

Levidis, Anastasios (1892). “*Pragmatia peri tis en Kapadokia lalumenis glosis ipo Anastasiu M. Levidu* [*A treatise by Anastasios M. Levidis on the language spoken in Cappadocia*]”. Manuscript deposited at the Center of Asia Minor Greek Studies, Athens.

doi Levin, Theodore and Omer Preminger (2015). “Case in Sakha: Are two modalities really necessary?” In: *Natural Language & Linguistic Theory 33*, pp. 231–250.

Lightfoot, David (1979). *Principles of diachronic syntax*. Cambridge: Cambridge University Press.

Lightfoot, David (1991). *How to set parameters: Arguments from language change*. Cambridge, MA: The MIT Press.

Lightfoot, David (1999). *The development of language: Acquisition, change and evolution*. Oxford: Blackwell.

doi Lightfoot, David (2006). *How new languages emerge*. Cambridge: Cambridge University Press.

doi Lightfoot, David (2017). "Acquisition and learnability". In: *The Cambridge handbook of historical syntax*. Ed. by Adam Ledgeway and Ian Roberts. Cambridge, UK: Cambridge University Press, pp. 381–400.

Marantz, Alec (1991). "Case and licensing". In: *Proceedings of the 8th Eastern states conference on linguistics*. Ed. by Benjamin Westphal and Hee-Rahk Chae. Ithaca, NY: CLC Publications, pp. 234–253.

doi Muysken, Pieter (2010). "Scenarios for language contact". In: *The handbook of language contact*. Ed. by Raymond Hickey. Malden, MA: Wiley-Blackwell, pp. 265–281.

doi Næss, Åshild (2004). "What markedness marks: The markedness problem with direct objects". In: *Lingua 114*, pp. 1186–1212.

doi Niyogi, Partha (2006). *The computational nature of language learning and evolution*. Cambridge, MA: The MIT Press.

Niyogi, Partha and Robert Berwick (1997). "A dynamical systems model for language change". In: *Complex Systems 11*, pp. 161–204.

doi Pinker, Steven and Alan Prince (1988). "On language and connectionism: Analysis of a parallel distributed processing model of language acquisition". In: *Cognition 28*, pp. 73–193.

Pintzuk, Susan (1999). *Phrase structures in competition: Variation and change in Old English word order*. New York: Garland.

doi Qi, Peng et al. (2020). "Stanza: A Python natural language processing toolkit for many human languages". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. url: https://nlp.stanford.edu/pubs/qi2020stanza.pdf.

Ralli, Angela (2000). "A feature-based analysis of Greek nominal inflection". In: *Glossologia 11–12*, pp. 201–227.

Roberts, Ian (2021). *Diachronic syntax*. Second edition. Oxford, UK; New York, NY: Oxford University Press.

doi Roberts, Ian and Anna Roussou (2003). *Syntactic change: A Minimalist approach to grammaticalization*. Cambridge, MA: Cambridge University Press.

Sarantidis, Archelaos I. (1899). *I Sinasos* [*Sinasos*]. Athens: Tipografion Ioannou Nikolaïdu.

Seliger, Herbert (1996). "Primary language attrition in the context of bilingualism". In: *Handbook of second language acquisition*. Ed. by William C. Ritchie and Tej K. Bhatia. New York: Academic Press, pp. 605–626.

Silverstein, Michael (1976). "Hierarchy of features and ergativity". In: *Grammatical categories in Australian languages*. Ed. by R. M. W. Dixon. Canberra: Australian Institute of Aboriginal Studies, pp. 112–171.

doi Spyropoulos, Vassilios (2020). "Abstract and morphological case in a nominative-accusative system with differential case marking". In: *Case, agreement, and their interactions: New perspectives on differential argument marking*. Ed. by András Bárány and Laura Kalin. Berlin: De Gruyter, pp. 175–218.

Spyropoulos, Vassilios and Maria-Anna Tiliopoulou (2006). "Definiteness and case in Cappadocian Greek". In: *Proceedings of the 2nd International Conference on Modern Greek Dialects and Linguistic Theory*. Ed. by Mark Janse, Brian D. Joseph, and Angela Ralli. University of Patras, pp. 365–378.

Theodoridis, Theodoros (1939). “Diorthosi kimenu ke scholia paramithion pharasiotikon [Corrections and comentary on the Pharasiot Greek stories collected by R. M. Dawkins]”. Unpublished Manuscript.

Theodoridis, Theodoros (1960). “Pharasiotikes paradosis, mithi kai paramithia [Pharasiot customs, myths and stories]”. In: *Laografia: Deltion tis Ellinikis Laografikis Etaireias 19*, pp. 221–263.

Theodoridis, Theodoros (1964). “Pharasiotikes paradosis, mithi kai paramithia 2 [Pharasiot customs, myths and stories 2]”. In: *Laografia: Deltion tis Ellinikis Laografikis Etaireias 21*, pp. 209–336.

Theodoridis, Theodoros (1966). “Pharasiotikos istorikos dialogos [A historical dialogue in Pharasiot Greek]”. Unpublished Manuscript.

Theodoridis, Theodoros. (undated). “Dictionary of the Greek dialect of Pharasa”. Unpublished Manuscript, probably completed in 1950s.

Triantafyllidis, Manolis (1938). *Neoeliniki grammatiki I: Istoriki isagogi* [*Modern Greek grammar I: A historical introduction*]. Thessaloniki: Institute for Modern Greek Studies.

doi van Coetsem, Frans (1988). *Loan phonology and the two transfer types in language contact.* Dordrecht: Foris.

von Heusinger, Klaus and Jaklin Kornfilt (2005). “The case of the direct object in Turkish: Semantics, syntax and morphology”. In: *Turkic languages 9*, pp. 3–44.

Vries, Mark de (2002). *The syntax of relativization*. Utrecht: LOT.

Weinreich, Uriel (1953). *Languages in contact. Findings and problems*. The Hague: Mouton.

Weinreich, Uriel, William Labov, and Marvin Herzog (1968). “Empirical foundations for a theory of language change”. In: *Directions for historical linguistics: A Symposium*. Ed. by Winfred P. Lehmann and Yakov Malkiel. Austin, TX: University of Texas Press, pp. 95–188.

Winford, Donald (2003). *An introduction to contact linguistics*. Malden, MA: Blackwell.

doi Winford, Donald (2005). “Contact-induced changes. Classification and processes”. In: *Diachronica 22*, pp. 373–427.

doi Woolford, Ellen (2006). “Lexical case and inherent case and argument structure”. In: *Linguistic Inquiry 37*, pp. 111–130.

doi Yang, Charles (2002). “Grammar competition and language change”. In: *Syntactic effects of morphological change*. Ed. by David Lightfoot. Oxford: Oxford University Press, pp. 367–380.

doi Yang, Charles (2016). *The price of linguistic productivity: How children learn to break the rules of language*. Cambridge, MA: The MIT press.

Yang, Charles (2018). *A User’s Guide to Tolerance Principle*. http://ling.auf.net/lingbuzz/004146. Accessed: 2022-12-03.

Address for correspondence

Ümit Atlamaz
Dept. of Linguistics
Boğaziçi University
John Freely Hall 307
34342 Bebek, Istanbul
Turkey
umit.atlamaz@bogazici.edu.tr
https://orcid.org/0000-0003-1657-9654

Co-author information

Metin Bagriacik
Dept. of Linguistics
Boğaziçi University
metin.bagriacik@bogazici.edu.tr
https://orcid.org/0000-0003-4544-7503

Publication history

Date received: 16 March 2023
Date accepted: 4 April 2024
Published online: 17 May 2024