The acquisition of Hindi split-ergativity and Differential Object Marking by Dutch L1 speakers: systematicity and variation

Aaricia Ponnet[1] and Ludovic De Cuypere[2]

[1] Ghent University

Department of Linguistics / Languages and Cultures INDIA

MULTIPLES / SANGH

[2] Ghent University

Department of Linguistics (GLIMS)

Vrije Universiteit Brussel (BCLS)

Contact: Aaricia.Ponnet@ugent.be

Abstract

We investigated the acquisition of Hindi split ergativity (zero or ne-marking) and Hindi Differential Object Marking (zero or ko-marking) by L1 speakers of Dutch. Both grammatical

phenomena are conditioned by multiple syntactic and semantic features. On a descriptive level, the study aims to examine when and how Dutch-speaking learners acquire and apply the conditional features associated with ne- and ko-marking in Hindi as a foreign language (HFL). A specific learner corpus was created based on a picture description task that elicited semi-spontaneous oral production data from 15 Dutch-speaking learners of Hindi, from four cross-sectional stages of the Hindi course trajectory. We annotated the corpus data for multiple features associated with ne- and ko-marking. Using a mixed-effects logistic regression analysis, we found an increase in the use and accuracy of each case marker over the different years of study, but individual learner profile analyses revealed considerable intersubject differences in learner behaviour. We argue that the developmental stages for the emergence of ne- and ko-marking are in line with predictions based on Processability Theory (Pienemann 1998). We additionally include mastery level analysis to account for a combined perspective on language development (Hulstijn 2015). Our findings reveal that that HFL learners reach higher mastery levels for split ergativity than for DOM, even though DOM (ko-marking) emerges before split ergativity (ne-marking). We conclude that developmental stages and between-learner variation are not mutually exclusive.

Funding

This work was supported by the Fonds Wetenschappelijk Onderzoek under Grant number 11G8321N.

Acknowledgements: to add…

This work was supported by the [Funding Agency] under Grant [number xxxx].

Data Availability Statement: The authors confirm that the data supporting the findings of this study and the replication code in R are available through TROLLing: Ponnet, Aaricia; De Cuypere, Ludovic, 2023, "Replication Data for: The acquisition of Hindi split-ergativity and Differential Object Marking by Dutch L1 speakers: systematicity and variation", https://doi.org/10.18710/3YWQ8R; DataverseNO, V1.

1. Introduction

Over the past two decades, an increasing number of linguists have advocated for a focus on structural differences between languages and for "placing diversity at centre stage" (Evans & Levinson, 2009, p. 429). In the field of second language acquisition (SLA) particularly, Stoll and Bickel (2013) observe that SLA research would benefit from broadening its scope to typologically diverse languages. Especially in studies that follow a longitudinal design, the most common L2 under study remains English (71%, McManus 2022). Fortunately, research on language acquisition has been more inclusive with research on minority languages (e.g. Cenoz & Gorter 2019), heritage languages (e.g. Mardale & Montrul 2020), learners from different SES backgrounds (e.g. Douglas Fir Group 2016), and previously neglected majority languages in multilingual communities (e.g. Heugh 2013, García 2011). This paper aims to contribute to this shift by focusing on an under-researched yet widely spoken language, namely Hindi[1] and investigates the acquisition of Hindi as a Foreign Language (HFL).

Hindi is currently one of the world's most spoken languages (Eberhard, Simons, & Fennig, 2021). However, empirical research on Hindi, let alone its acquisition, is scarce. The present study has two complementary goals. First, on a descriptive level, we examine how Hindi split ergativity and DOM are acquired by L1 speakers of Dutch (Dutch HFL), a typologically distant language which lacks both ergativity and DOM. The study thus answers the call for a broader scope in foreign language acquisition research, i.e. the inclusion of languages that have hitherto been overlooked or very rarely been researched within the field of SLA, and, for morpho-syntax, the inclusion of languages that portray different case and alignment systems. Secondly, the study aims to contribute to the discussion about whether the foreign language acquisition process follows clearly defined developmental stages or whether the between-learner variation is too substantial to assume such stages.

Unlike most Indo-European languages, which have (remnants of) nominative-accusative alignment, Hindi has partly ergative-absolutive alignment. Hence, Hindi overtly marks for case. The Hindi case system is based on postpositional marking whereby the noun is followed by a postposition/marker. Bare nouns and pronouns occur in the nominative by default, whereas several postpositions, such as *-ne* and *-ko*, indicate the case role of the arguments (Kachru, 2006). In Hindi, the subject and the direct object (DO) can be either null-marked or overtly marked for case depending on certain conditions. These types of case marking, which are attested in other languages as well, are known as split ergative case marking of the subject, and Differential Object Marking (DOM). We will regularly use the names of the Hindi case markers under study and refer to split ergativity in Hindi as *-ne* marking and to DOM as *-ko* marking. Both are conditioned by syntactic and semantic features, which we will describe in full detail with the necessary examples in the following section.

Our study is theoretically informed by Processability Theory (PT) (Pienemann 1998), a theory which strongly supports a developmental stage approach to foreign language acquisition. According to PT, lexical entries are annotated for syntactic and semantic features, but these features will only gradually become available through several processing procedures and the unification of features within these procedures. In other words, the learners are able to form sentences from an early stage onwards, but it is only at later stages that information exchange between the different constituents in a sentence becomes functional. The latter is crucial for a correct use of *-ne* and *-ko* marking.

We start from PT's assumption that the language acquisition trajectory follows certain developmental stages. Previous PT-studies on case marking (e.g., Baten 2013, Artoni 2013) have proposed a stage of positional marking, where constituents in canonical subject or object position would be marked with the corresponding case (but errors would occur when constituents were not in the canonical position). However, for Dutch learners of Hindi such a stage seems unlikely because Hindi has no opposition between preverbal and postverbal arguments, relatively free word order and differential case marking. As an alternative to positional marking, we hypothesize that Dutch HFL learners go through a stage of semantic mapping before acquiring functional marking, whereby the case marker will be mapped onto one of the features that determines *-ne* and *-ko* marking (without a functional distinction between these features). *-ne* marking would accordingly first be linked to agentive-like arguments (i.e. transitive subjects), while *-ko* marking would first be associated with the dative and goal object. *-ko* marking of the direct object would only emerge later as they assume an initial association of one form with one function – however, at this later stage, learners hypothesize that *-ko* as a DO-marker will be mapped onto specific arguments first. A more detailed developmental process is outlined in section 2 (table 2).

PT adopts Levelt's (1989) Model of Language Production (which is based on incremental language generation), and assumes a developmental trajectory that follows stages according to the processability of a certain linguistic feature, which implies that learners will acquire some features before others. The existence of developmental stages has been challenged by a.o. Hulstijn (2015), who claims that language acquisition develops beyond emergence and SLA research should also be able to account for between-learner variation and murky production data. More recently, PT has started to include variational approaches to language acquisition (mainly Dynamic Systems Theory (DST) (a.o. Verspoor et al. 2008) to examine whether variation is still compatible with a developmental framework. The latter is argued, for instance, by Dyson & Håkansson (2017) from a PT perspective. Moreover, DST studies within SLA have been focussing more on patterns of variation, and have defined several types of interaction that could account for the development of variation in learner language (e.g., Lowie & Verspoor 2015; Verspoor et al. 2021). These interactions can be competitive (Verspoor, Lowie & Dijk, 2008), when the acquisition of one feature prevents the acquisition of another, supportive (Spoelman & Verspoor, 2010), i.e. features enhance each other's acquisition, or a precursor (Caspi, 2010) for other developments, i.e. the acquisition of one feature enables the acquisition of another.

Hindi split-ergativity and DOM are complex features to acquire by L1 Dutch learners, as these features are not present in Dutch. We thus expect to find considerable variation amongst these learners. Moreover, given that learners need to acquire different case markers which potentially interact, we aim to verify whether and how the variation between the learners is compatible with the developmental stages that we hypothesize. More specifically, we aim to examine whether the developmental stages for both *-ne* and *-ko* marking coincide or not. We therefore follow Hulstijn's (2015) call for a combined perspective and analyse emergence as well as mastery levels of the case markers under study. We hypothesize that developmental stages and variation are not mutually exclusive and that a learner may have reached a higher mastery level for *-ne* than *-ko* marking or vice versa, even though both of the case markers have emerged.

To test our hypotheses, we used a picture description task to elicit semi-spontaneous oral production data from 15 learners of Hindi at a Belgian institute. Following a pseudo-longitudinal design, we gathered data on one occasion from students in either their first, second, third, or fourth year of the

Hindi course trajectory. We annotated the corpus data for multiple features related to *-ne* and *-ko* marking: Perfectivity, Transitivity, Animacy, and Semantic Role. Using mixed-effects logistic regression analysis, we modelled how the learners used *-ne* and *-ko* marking in relation to the features. Given that we have (cross-sectional) data from multiple years and that we include Participant as a random effect, we are able to take between-learner variation into account. We additionally offer a more qualitative interpretation of each learner's linguistic data in order to create individual learner profiles.

The article is organized as follows. Section 2 describes the case marking system of Hindi, focusing on split ergativity and DOM. The section further elaborates on the research questions of the article and develops the research hypotheses of the study, focusing on the hypothesized developmental stages and variation among learners. Section 3 outlines the theoretical framework, while section 4 provides more details on the background of the learners, outlines the collected corpus data, and describes the data analyses. Sections 5 and 6 present the results for *-ne* marking and *-ko* marking, respectively. Section 7 discusses the results and the theoretical implications of our findings. Section 8 draws general conclusions and makes suggestions for future research.

2. Hindi as a Foreign Language

In Hindi, the unmarked noun corresponds to the nominative case, the only case in Hindi which is "phonologically null" (Butt, 1993, p. 92). Other cases are indicated by a postposition/marker following the noun. The noun takes the oblique when followed by a postposition/marker, e.g. the noun *laṛk-ā* 'boy' will take the oblique *-e* ending when followed by *ke liye* 'for', as in example (1):

(1) *Laṛk-e=ke liye*

boy.OBL=GEN for

'For the boy'

Note that certain nouns, such as singular feminine nouns (e.g. *laṛkī* 'girl', *aurat* 'woman'), have the same form for the nominative case and the oblique case. We will refer to this form as 'invariable'.

The present study focusses on two case markers: (i) *-ne*, which marks ergative case and (ii) *-ko*, which marks the Direct Object (DO) as well as other semantic roles such as the indirect object (IO)/recipient, experiencer constructions and the goal object.

2.1 -ne marking: Split ergativity

In a language such as Hindi, an opposition is made between the transitive subject (A) on the one hand, and the intransitive subject (S) and the patient (O) on the other hand. The marked agent is said to be in the ergative case, the intransitive subject and the patient remain in the unmarked nominative case. This is called an Ergative-Nominative system[2]. Typological research has shown that fully ergative languages

are rare. Most ergative languages feature a split in the system. In Hindi, that split resides on the border of perfective and non-perfective verbs (Verbeke, 2013). As illustrated in example (2), the agent of a transitive perfective verb is marked with the ergative case marker *-ne*; in example (3), the single argument of the intransitive perfective verb remains unmarked. All subjects of non-perfective verbs (such as imperfectives, progressives, etc.) will also remain unmarked, as in (4).

(2) *batman=**ne*** *laṛkī=kī* *jān.ø* *bacāyī*

batman=ERG girl=GEN life.NOM rescue-PERF

'Batman saved the girl's life'

(3) *batman.ø* *tairā*

batman.NOM swim-PERF

'Batman swam'

(4) *batman.ø* *laṛkī=kī* *jān.ø* *bacātā + hai*

Batman.NOM girl=GEN life.NOM rescue-IMPERF + AUX

'Batman saves the girl's life.'

2.2 -ko marking: one marker, multiple roles

-Ko marks multiple roles in Hindi. Firstly, *-ko* marks the DO/patient in Hindi. Hindi has Differential Object Marking (DOM), like many other languages such as Spanish (Montrul & Sánchez-Walker, 2013), Romanian (Mardale, 2008) and Turkish (Haznedar, 2006). Generally speaking, DOM is motivated by DO features such as animacy and definiteness/specificity, but the saliency of the features differs cross-linguistically (Malchukov, 2008). In Hindi, DOM is related to the factors animacy and definiteness/specificity (Kachru 2006, Butt 1993, Klein & de Swart 2011, Aissen 2003, Mohanan 1994). For methodological purposes, we will only refer to the feature specificity from now on (see below). Examples (5 to 7) taken from Mohanan (1994: 80)[3] illustrate the DOM pattern in Hindi:

(5) *ilā=ne* ***bacc-e=ko*** *uṭhāyā*

Ila=ERG child-OBL=OBJ lift-PERF

'Ila lifted a/the child'

(6) *ilā=ne* ***hār=ko*** *uṭhāyā*

Ila=ERG necklace=OBJ lift-PERF

'Ila lifted the/*a necklace'

(7) *ilā=ne* ***hār***.ø *uṭhāyā*

Ila=ERG necklace lift-PERF

'Ila lifted a/the necklace'

According to Mohanan's account, the rule for animate DOs (like 'child' in (5)) seems to be quite straightforward: animate DOs are obligatorily marked with *-ko*. For inanimate nouns (like 'necklace' in (6) and (7)) the role of specificity is crucial to the marking with *-ko*: a *-ko* marked DO will always be specific. A non *-ko* marked DO can be either specific or non-specific. However, if the speaker intends a non-specific meaning, there is no optionality and *-ko* should be absent.[4]

Mohanan (1994: 80) gives an example in Hindi were the non-human animate DO remains unmarked when it is non-specific, such as in the sentence *ravī (ek) gāy kharīdnā cāhtā hai* 'Ravi wants to buy a cow (with no particular cow in mind)'. These instances where non-specific animate DOs remain unmarked have been subject to typological discussion on Hindi DOM and linguists do not seem to agree whether animacy outweighs specificity or vice versa (a.o. Butt 1993; Mohanan 1994; Montaut, 2004; Malchukov 2008; López 2012). In their paper on Hindi DOM, Verbeke & Ponnet (2022) argue that in those instances where an animate noun remains unmarked, the role of specificity can explain the exception. This does not imply, however, that specificity outweighs animacy. Moreover, in Hindi textbooks these exceptions are usually not touched upon.

Ko also obligatorily marks the recipient, as in (8), and several other functions (experiencer subject, desiderative, obligative) which we have grouped under the term 'experiencer construction' (Butt 2006).

(8)	*Ritā*	*Sitā=**ko***	*akhbār*	*degī*
	Rita.NOM.SG	Sita=OBJ	newspaper=NOM.SG	give.FUT.3SG

"Rita will give (a/the) newspaper to Sita."

EXPERIENCER SUBJECT

(9) *nādyā=**ko*** *ḍar* *lagā*

nādyā=DAT fear.NOM be attached-PERF

'Nadya was afraid'

DESIDERATIVE

(10)*nādyā=**ko*** *ek* *kitāb* *cāhiye*

nādyā=DAT one book.NOM want

'Nadya wants a book.'

OBLIGATIVE

(11)*nādyā=**ko*** *Lucknow* *jānā* *hogā*

nādyā=DAT Lucknow.ACC go-INF be-FUT

'Nadya will have to go to Lucknow.'

Finally, *-ko* can mark goal objects as in (12) (Everaert, 2004, p. 28) but here, too, *-ko* is more than often dropped, resulting in a bare locative as in (13) (Butt & King, 2004, p. 168). Remember that, usually, the Hindi noun takes the oblique when followed by a postposition/marker. When *-ko* is dropped with goal objects such as *dākkhāne*, the noun has the oblique *-e* ending even though there is no overt case marking:

(12) *Maiṃ* *śahar=ko* *jātī + hūṃ*

I.NOM city=ACC go-IMP.F.1SG

'I go to the city'

(13) *Adnān* *ḍākkhān-e* *gayā + hai*

Adnan.M.Sg.Nom post office.M.Sg.Obl go-Perf.M.Sg + be.Pres.3.Sg

'Adnan has gone to the post office.'

2.3 Previous studies

Studies exploring the acquisition of case marking in Hindi initially focused on Hindi as a first language (L1) (Narasimhan et al. 2005, Narasimhan 2005). Note that we follow the common practice in SLA studies to use the term L1 for the native language(s) of a speaker, and the term L2 for any language that has been acquired after the L1(s). Recently there has also been a gradual rise in studies with adult speakers (Montrul et al. 2012, Baten & Verbeke 2015, Ponnet et al. 2016, Ranjan 2016, Montrul et al. 2019a). The latter studies consider different learner populations and different linguistic areas. Montrul et al. (2012) examined the acquisition (and possible loss) of case and agreement among heritage speakers of Hindi in the US. Four studies have been performed with learners of Hindi as a Foreign Language, two on the development of the ergative case marker (Baten & Verbeke, 2015; Ranjan, 2016), one on the development of DOM (Ponnet et al. 2016), and one on both case markers (Montrul et al. 2019a). All of these studies focussed on oral language production. Additionally, Ranjan (2016) and Montrul et al. (2019a) tested the HFL learners' receptive skills using, a. o., an acceptability judgment task. Regardless of the learner type, all of the studies reported omission of *-ne* marking and *-ko* marking. It seems that case marking in Hindi is particularly difficult to acquire, and the different studies provided different explanations as to why this is so. We address the findings of these studies, and in particular their implications for our current study, during the discussion.

3. Theoretical background

Recent studies on HFL (Baten & Verbeke 2015; Ponnet et al. 2016; Ranjan 2016; Montrul et al. 2019a) have shown that case marking poses specific problems for FL learners. One explanation is that learners cannot solely rely on a one-to-one mapping between form and function but have to take into account multiple factors simultaneously. Whereas ergative case marking (*-ne* marking) of the subject/agent depends on the transitivity and perfectivity of the verb (volitionality has also been implicated, see among others Butt (2017)), objective case marking (*-ko* marking) depends on the animacy and specificity of the DO/patient. In addition, *-ko* also marks the IO/recipient, goal object and the experiencer.

Table 1: *-ne* marking and *-ko* marking in Hindi

***-ne* marking (ERG)** of the subject/agent	*Ø / -ne*	***-ko* marking (DOM)** of the DO/patient	*Ø / -ko*
Imperfective		Animate	
Transitive	Ø	Specific	*-ko*
Intransitive	Ø	Non-specific	*-ko* [/Ø]
Perfective		Inanimate	
Transitive	*-ne*	Specific	*Ø/-ko*
Intransitive	Ø	Non-specific	Ø
		-ko marking (other roles)	*Ø / -ko*
		Recipient	*-ko*
		Experiencer	*-ko*
		Goal	*Ø / -ko*

Some analyses of Hindi consider the *-ko* marker in instances with an animate non-specific DO as optional (see Verbeke & Ponnet (2022) for an overview of the existing studies on Hindi DOM). It suffices to say that there is no consensus on the matter, and some linguists even argue that *-ko* is grammaticalising towards an accusative case marker (Rajesh Bhatt, personal communication, 19 June, 2018). Up to today the most detailed description of the use of *-ko* as a DO marker remains that of Mohanan (1994), which is the description we follow for the present study.

The interaction between grammatical function and semantic role in this type of case system challenges theories of second language acquisition that have focused on developmental stages (Dyson & Håkansson, 2017), as these have previously mainly focused on nominative-accusative languages with a one-on-one relationship between the grammatical function and its case marker. For these languages (see a.o. Baten (2013) on L2 German, and Artoni & Magnani (2015) on L2 Russian) the following stages have been proposed: all- nominative > direct mapping > positional marking > functional marking. Direct mapping for these languages implies an opposition between a preverbal nominative argument and a postverbal accusative argument. Positional marking means that case forms are mapped onto an argument position (but errors occur when the argument does not appear in its canonical position). Only at the functional marking stage will learners be able to deviate from the canonical word SVO word order and will the case forms be assigned to the correct grammatical function.

For this study, our theoretical approach focusses on developmental sequences as well as variation by drawing upon theoretical frameworks such as Lexical Functional Grammar (LFG) (Bresnan et al., 2015; Butt, 2006), Processability Theory (PT) (Pienemann, 1998) and Dynamics Systems Theory (DST) (Larsen-Freeman, 2007). We propose an acquisition trajectory that envisages certain developmental sequences. More specifically, we build on Baten & Ponnet (2023) who propose several developmental stages via grammatical phenomena, as described by LFG, such as feature unification and mapping of arguments.

As Hindi is a "role dominated language" (Montaut, 2004, p. 211), Baten & Ponnet (2023) focus on semantic roles as well as grammatical functions and propose a mapping of arguments through what they call 'semantic mapping'. 'Semantic mapping' predicts that the FL learners will map the case markers onto the grammatical function by focusing on one of the semantic features that guide the use of these markers.

We similarly assume that the learners' language behaviour will be influenced by 'relative prominence' (Mohanan, 1994) in the thematic hierarchy (agent > goal > patient/theme > locative), but also by proto-role entailments that deconstruct roles such as 'agent' into more basic components such as volition, sentience, causation, etc. (Dowty, 1991). We accordingly predict that, initially, the learners will not be able to produce complex morphological utterances, as in there will only be utterances of acquired words (Stage 1 in Table 2), e.g. rote-learned formulaic sentences.

Table 2: developmental trajectory

Stage	**Processing procedures**	**Case in Hindi**
4	S-procedure	Functional nominative marking *-ne* marking agents of transitive perfectives *-ko* marking of animate DO *-ko* marking of non-human animate DO

3	Phrasal procedure	*-ko* marking specific DO
2	Category procedure	*-ne* marking of agentive-like arguments *-ko* marking of recipient *-ko* marking of experiencer
1	Word/lemma access	Default null-marking

The learners might produce multi-word sentences but we expect the nouns to be null-marked by default, i.e. the nominative case form, but with no opposition of any other core case marker. Because of the latter, we do not assume the presence of functional case in this stage. In the category procedure, in stage 2 that is, an association between *-ko* marking and the recipient/goal object arises. The authors motivate this because they assume that *-ko* will firstly be associated with one function, and because there is a one-on-one relationship between the semantic role of recipient and (dative) *-ko* marking. Furthermore, they assume that *-ne* marking will be associated with agentivity, because of the relative prominence of this semantic role, as well its proto-role entailment of volitionality, making *-ne* marking possible in this stage. To summarise, at this stage the case markers will be directly mapped onto the associated semantic feature of the constituents. There is no phrasal information exchange. As Baten & Ponnet (2023) assume that *-ko* will be mapped onto one function initially, an association with the DO is only assumed at the next stage, the inter-phrasal procedure (stage 3). Following Verbeke & Ponnet's (2022) analysis of Hindi DOM, in which the feature specificity explains sentences in which the animate DO remains unmarked, Baten & Ponnet (2023) propose an association with specificity at this stage. *-ko* marking will, according to their account, be triggered by the earlier use of *-ko* on goal objects. Note that at this stage, PT considers all case forms to be underspecified for grammatical function – we thus only assume semantic mapping of the case forms but no functional case marking. Further down the road, during the S-procedure (stage 4), information exchange between the verb phrase on transitivity and aspect and the subject noun phrase will emerge, making appropriate ergative case marking possible (i.e. the development of split ergativity). The information on the animacy of the patient will become available and a functional opposition will emerge between nominative null-marking on the one hand, and functional *-ne* and *-ko* marking on the other. In addition to Baten & Ponnet (2023), we propose an emergence of experiencer *-ko* marking at stage 2, where we expect chunking, i.e. the experiencer will be used in the same construction, as well as semantic mapping. At stage 4, we additionally make a distinction between human animate patients and non-human animate patients (following Montrul et al. 2019a).

It is important to know that PT makes claims only on the emergence of certain features, and that the development from emergence to mastery is currently being explored in PT studies by including a variationist approach. Dyson (2021), for example, proposes a dynamic variation approach, in which different routes of grammatical development are predicted within the developmental stages. She theorizes that it is possible to predict these routes of development based on the inherent features of the language acquired and shared elements between the L1 and the LX.

Variation in second language acquisition has been extensively discussed in studies that use the framework of Dynamic Systems Theory (Larsen-Freeman 2007). In this account, variation is regarded as an inherent feature of language acquisition, i.e. it is an indicator of change and thus essential for the development of the interlanguage of the learner. Of course, the interlanguage of the learner is not a

monolith; different parts of the interlanguage develop at different moments in time at a different pace. A handful of studies describes several kinds of interactions between these developing systems: competitive (Verspoor, Lowie, & Van Dijk, 2008), supportive (Spoelman & Verspoor, 2010), and precursor (Caspi, 2010). Applied to the acquisition of Hindi case marking, this implies that the developmental routes of the acquisition of the different case markers can cause different types of relationships between these features. Accordingly, we expect that one emerging case marker could be the precursor for the emergence of another. Thus, the emergence of the ergative case marker could result in a competitive interaction with verbal agreement whereas the emergence of non-DO *-ko* marking could result in a supportive interaction for the emergence of *-ko* as a DO-marker.

4. Methodology

PT is a framework with a strong focus on the individual language learner. In accordance with PT and with current methodology of multiple case study analysis in SLA (which aims at fine-grained analysis of smaller populations rather than analysing large sets of randomised data), we used a picture description task to elicit semi-spontaneous oral production data from N = 15 learners of Hindi at a Belgian institute. Following a pseudo-longitudinal design, we gathered data on one occasion from students in either their first, second, third, or fourth year of the Hindi course trajectory.

Participants

We recruited 15 Dutch-speaking learners of Hindi on a voluntary basis. As randomised sampling was not feasible for this population (HFL learners are scattered all over Europe in small numbers) and we wanted to be able to control for their L1 and their learning environment, we decided to work with a convenience sample. The students were informed that they were partaking in a study on Hindi language development, but the specific focus of the study, viz. the acquisition of the Hindi case and agreement system, was not disclosed. The students were enrolled in a Hindi language course at a university in Belgium. All students were enrolled in a full-time academic program and belonged to similar SES-backgrounds.[5] Table 3 provides an overview of the participants, with their L1, gender, year of study, and the amount of instruction they received.

Table 3: Background participants

Year	Number of participants	Female / Male	Hindi Instruction/week
1	3	1 / 2	4h30
2	3	3 / 0	5h
3	6	4 / 2	6h
4	3	2 / 1	4-5h

Ten participants were female, five were male. The participants' age ranged beween 18-35, with the median age at 20. Their learning environment was instructed. The participants were interviewed at the end of the first semester of the academic year. This means that the learners of year 1 had had 1 semester of instruction, the learners of year 2 had had 3 semesters of instruction, and the learners of

year 3 had had 5 semesters of instruction. The learners of year 4 had had 6 semesters of instruction, as in the first semester of year 4 they were studying Hindi texts but no longer received Hindi language instruction. All the students were enrolled in a full-time university education, with Hindi as their main subject. They received different amounts of instruction depending on their year of study. The teaching staff consisted of non-native speakers as well as a native visiting teacher. Participants from years 3 and 4 had spent at least 1 semester studying Hindi in India (study abroad semester). All of the students were acquainted with the *-ne* marker and the *-ko* marker on a comprehensive level by the end of year 1. This means that they were introduced to the entire Hindi grammar by the end of year 1, but received more instruction and exercises on the production of these markers during year 2.

All of the students had Dutch as their L1, and had already learnt French and English as additional languages before starting their Hindi course. They had no exposure to Hindi nor any other modern South Asian language prior to their Hindi language course[6]. These students were simultaneously enrolled in a Sanskrit language course. Among Hindi teachers the general assumption is that knowledge of Sanskrit is of great value to the students' vocabulary knowledge (although this has, so far, not been empirically tested), but as the grammar of Sanskrit, a classical language, differs greatly from the grammar of Hindi, a modern language, we do not expect any influence/transfer there.

Materials and procedure

The design of our study was based on earlier work by Baten (2013) and Baten & Verbeke (2015). We created our own picture description task to elicit spoken data from the Hindi learners in a semi-structured format, which we had already used successfully in an earlier study on Hindi case acquisition with a study abroad group (Ponnet et al. 2016). The participants were interviewed individually in a quiet classroom at the university, and presented with a task that contained different sets of pictures: 31 selected pictures from a comic book (Dulieu, 1974), one drawing with multiple characters and actions, as well as our own created material, a narrative of 33 pictures. The picture sets were ordered in a narrative structure, and the students were asked to describe in Hindi what they saw on these pictures in a story-like fashion. The task was self-paced, and aimed to test the learners' capacity to utilise their interlanguage grammar under the time constraints of spontaneous oral language production (Pienemann 1998). A brief vocabulary list was provided, consisting of nouns and verbs present in the picture description task. The pictures showed different contexts with a DO/patient argument and included (human and non-human) animate patients as well as inanimate patients.[7] In order to elicit non-specific patients, the pictures regularly introduced new patients in the narrative (e.g. by making a character point into the distance on one picture, and revealing what the character sees in the following picture). To test the use of the ergative case marker, the pictures in the narratives showed transitive as well as intransitive actions. The participants were asked to describe the narrative in a story-like fashion, for which the past is the required tense in Hindi. If they did not produce this tense spontaneously, they were explicitly asked to continue telling the story in the past tense. In this way, we elicited both perfective and non-perfective responses from each learner. To elicit specific DOs, the participants were asked to answer the question *āp kyā dekhte/dekhtī haiṃ* ('What do you see?') and were asked to produce sentences using a similar sentence:

Maiṃ	[…]	dekhtā/dekhtī+huṃ
I.NOM	[object seen on drawing]	to see.1SG.PRS

The aim was to create a more natural learner corpus, so we did not strictly control for all the contexts, but chose a semi-spontaneous elicitation design. We validated the task with two native speakers in India, to make sure the task elicited the contexts we wanted to investigate. In total, we collected 4h59m11s of data. Each interview lasted about 20 minutes, interview duration varied from 13m37s (min.) to 21m06s (med.) to 33m35s (max.).

The corpus data was transcribed manually in the word processing softwares Word/Pages and annotated in MS Excel by the first author of this study, who is an advanced non-native speaker of Hindi. A sample of the transcriptions was checked by a native speaker of Hindi, and a sample of the annotations by an advanced non-native speaker of Hindi other than the author of this study. The data is annotated for (i) Learner, (ii) Year of study of the learner, (iii) the use of *-ne* as an ergative marker, (iv) correct usage of the *-ne* marker, (v) the use of *-ko* as a Differential Object Marker, (vi) the use of *-ko* as another marker, and multiple features associated with *-ne* and *-ko* marking, including: (vii) specificity of the Direct Object, (viii) animacy of the Direct Object, (ix) transitivity of the sentence Verb, (x) perfectivity of the sentence Verb, (xi) other uses of the -ko marker, (xii) the semantic role of these other uses of the -ko marker. The annotated dataset is shared on TROLLing (see Data Availability Statement).

Statistical data analysis

We analysed the three types of case marking separately. First, we performed bivariate analyses and then we fitted (mixed-effects) regression models.

1. Ergative marker *-ne*: mixed-effect logistic regression model with an interaction between Year and Transitivity, and Participant as a random intercept. The bivariate analysis shows that learners predominantly use an ergative marker in perfective sentences (as is expected from the norm), so we fitted the model on the subset of perfective sentences only. Participant is included as a random intercept to account for clustering (learners create multiple utterances, so the observations are not independent).
2. DOM *-ko*: mixed-effects logistic regression model with an interaction between Year and DO animacy, and Participant as a random intercept. The bivariate analysis shows that *-ko* is never used with non-specific DOs, so we restricted the model to the subset with specific DOs.

3. Other uses of *-ko*: a logistic regression analysis with Year and Semantic role of other uses of *-ko* as main effects. We obtained a singular fit with a random intercept and therefore decided to drop this random effect. The interaction between Year and Semantic Role was not significant at the 5% significance level based on a LRT of nested models, so we also excluded this interaction from the model.

All analyses were performed with R (R Core Team, 2021) in RStudio (RStudio Team 2021). We used the following packages (in alphabetical order): dplyr (Wickham et al., 2021), effects (Fox, 2003), ggeffects (Lüdecke 2018), ggplot2 (Wickham, 2016), Hmisc (Harrell & Dupont, 2021), lattice (Sarkar, 2008), lme4 (Bates et al., 2015), multcompView (Graves et al. 2019), optimx (Nash, 2014; Nash & Varadhan, 2011), RColorBrewer (Neuwirth 2022), rms (Harrell 2021) sjPlot (Lüdecke 2022), The replication code for the analysis is shared in TROLLing (see Data Availability Statement).

Learner profile analysis

We additionally created individual learner profiles based on the utterances in the dataset. We mined the data for their behaviour concerning the different case markers in Hindi. We accordingly analysed the use of the case markers in relation to each other, i.e., whether a learner used *-ko* marking for the IO but not for the DO, or used *-ko* marking for all the roles but only with certain lemmas, or used *-ko* marking but not *-ne* marking and vice versa. Other behaviours were analysed as well, e.g., the portrayal of explicit knowledge of the Hindi grammar rules, hypercorrections, etc. The following section will show that learner behaviour and level of study do not necessarily coincide.

5. Results: *-ne* marking

The elicitation task produced a dataset of N = 1811 utterances, of which there are 1404 with a Subject that could be marked with *-ne*. Of these, 212 (12%) are found with the ergative marker *-ne*. Figure 1 shows the usage of the Ergative marker by year.

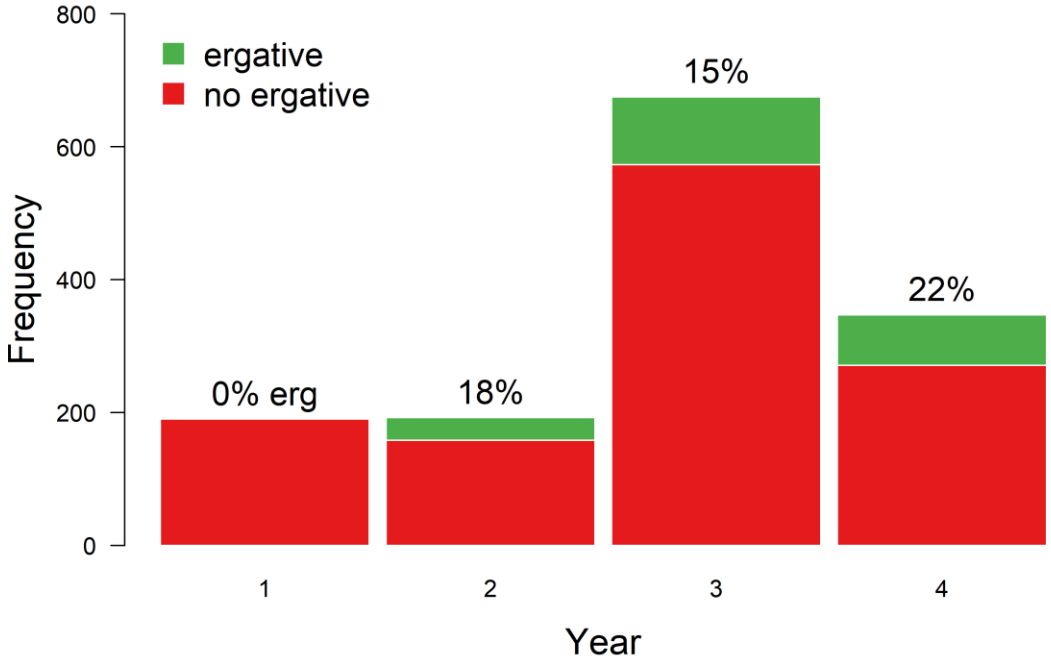

Figure 1: Usage of ergative marker by Year. Bar heights indicate absolute frequencies. Green refers to ergative uses, red to non-ergative uses. The proportion of ergative markers conditional on Year is added as a percentage.

No ergative marker was produced by first-year learners. We can therefore remove the first-year observation from further quantitative analysis.

Table 4 shows the relationship between Ergativity and Perfectivity over Year.

Table 4: Use of Ergative marker *-ne* in relation to Perfectivity by Year.

		Ergative marker	
	Perfective	**Yes**	**No**
Year 1	Yes	0	1
	No	0	181
Year 2	Yes	34	55
	No	0	92
Year 3	Yes	99	144
	No	3	397
Year 4	Yes	75	75
	No	1	208

There are only 4 cases where an ergative marker is used in a non-perfective sentence. This suggests that learners follow the rule that the ergative marker is only used in sentences with perfective aspect. This doesn't necessarily result in the correct use of ergative case marking, which becomes clear when we take transitivity into account.

Given that there is no variation in the use of the ergative marker in relation to perfectivity, we reduced the scope of our investigation to sentences with perfective aspect. Figure 2 shows the cross-sectional evolution by year of the use of the ergative marker in relation to transitivity:

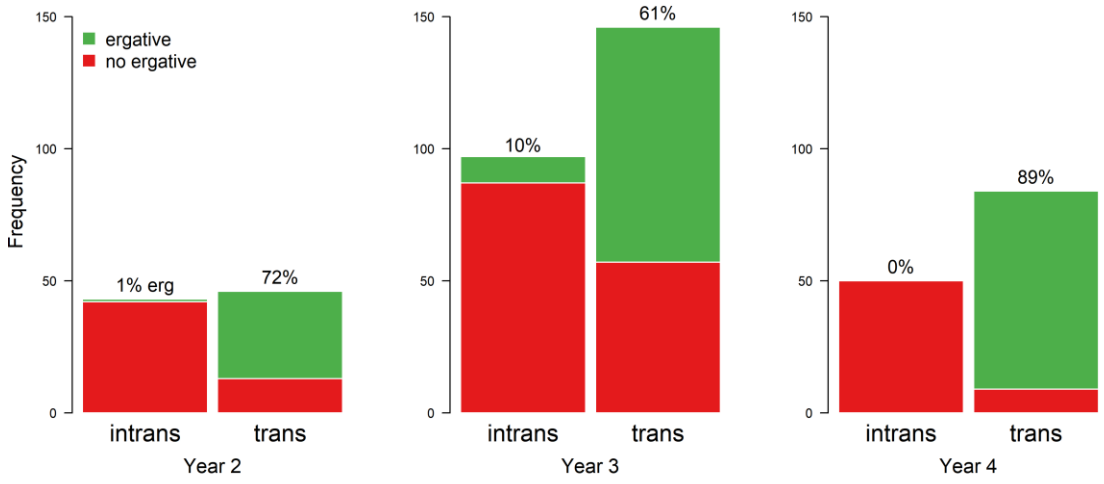

Figure 2: Use of ergative marker in relation to Transitivity by Year. Bar heights indicate absolute frequencies. Green refers to ergative uses, red to non-ergative uses. The proportion of ergative markers conditional on Transitivity by Year is added as a percentage. All sentences are perfective.

The ergative marker is nearly always used with transitive verbs and only rarely (11/208, 5%) with intransitive verbs. Nevertheless, a considerable number of subjects of transitive verbs are also used without the ergative marker (i.e. under-use of the ergative case marker). The latter is particularly the case in year 3, which suggests that the effect of Transitivity is stronger in year 2 and year 4.

Next, we present the results of a mixed-effects logistic regression with Ergativity as the outcome variable, an interaction effect between Transitivity and Year as a fixed factor, and with Learner as a Random intercept. We included an interaction effect to evaluate possible differences in the effect of Transitivity between the years (as suggested by the bivariate findings presented above). Given that no

ergative marker was used in Year 1 and only 4 ergative markers were used in imperfective sentences, we restrict our analysis to Years 2 to 4 and perfective sentences.

The interaction effect is significant at the 5% significance level, based on a likelihood ratio test of fitted models (G = 20, df = 2, P-Value < 0.001). Figure 3 visualizes this interaction effect. (Model estimates and other modelling details can be found in the replication code in the shared dataset).

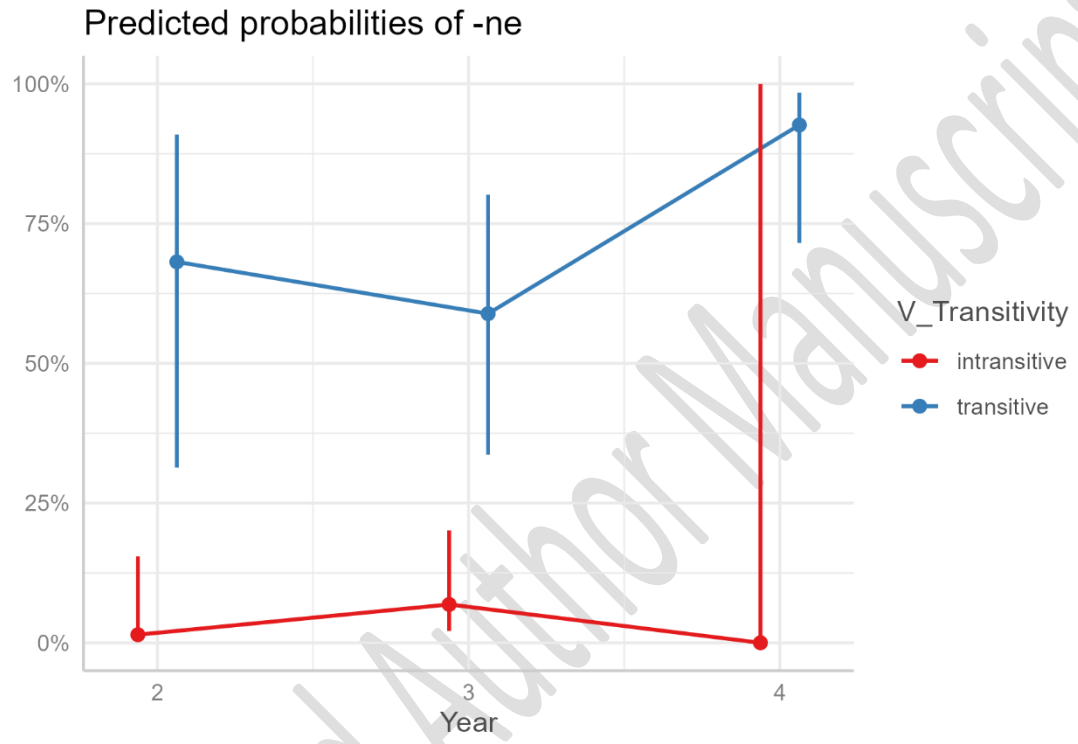

Figure 3: Effect plot for the interaction effect between Transitivity and Year on the usage of the Ergative marker *-ne*.

Overall, the ergative marker is most probable with transitive verbs. As regards the interaction effect, it appears that the difference between year 2 and 3 is rather marginal, while there is a very low probability for the ergative marker in year 4 with intransitive verbs, and a very high probability of the ergative marker with transitive verbs. The large confidence interval for intransitive in year 4 is because no intransitive sentence was produced with an ergative marker.

So over the years, a functional opposition appears to become established, and the results suggest that in year 4 the learners have acquired the rule that no ergative marker should be used in intransitive sentences. This is additionally confirmed by a post hoc analysis of the pairwise differences for Transitivity conditional on Year. Table 5 shows the pairwise difference estimates:

Table 5: Posthoc pairwise estimates for the difference between the predicted probabilities (Transitive – intransitive) conditional on Year. The predicted probabilities are visualized in Figure 3. SE = Standard error.

Year	Estimate (SE)	Z	P-Value
2	67%* (16%)	4.09	< 0.0001
3	52% (10%)	5.05	< 0.0001
4	93% (5%)	16.55	< 0.0001

* 67% is the difference in predicted probabilities between transitive and intransitive in Year 2, as visualised in Figure 3.

Individual learner effects are captured by the BLUPs (Best Linear Unbiased Predictors) visualized in Figure 4.

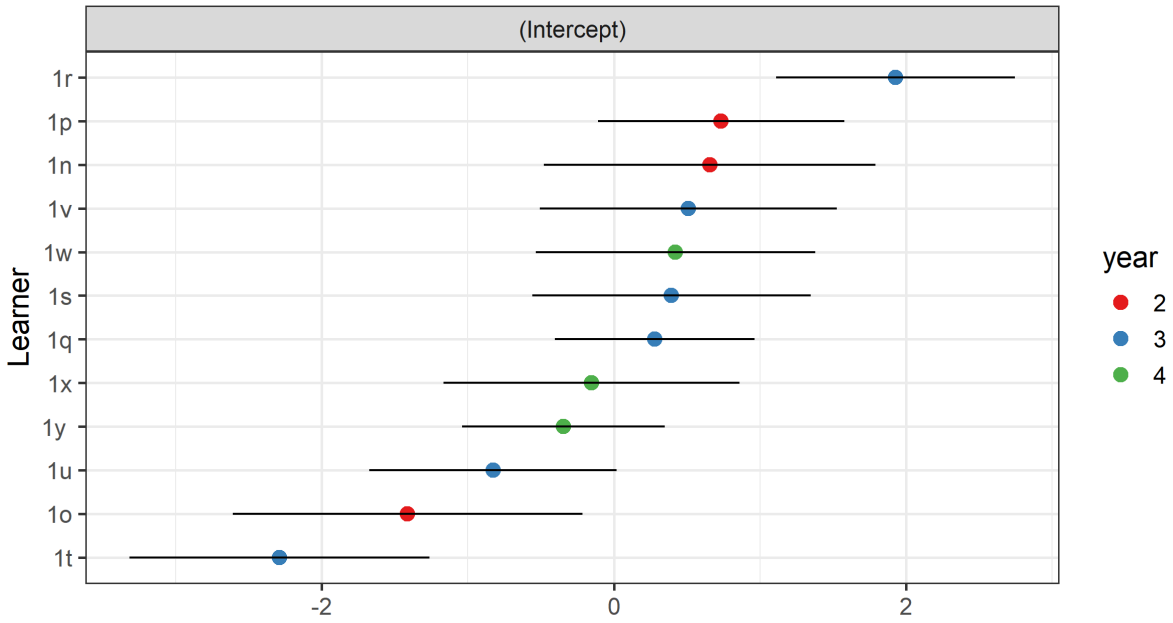

Figure 4: Best Unbiased Linear Predictors (BLUPS) for the random effect Learner.

In a random-effects model, the BLUPS account for the individual biases of each participant. More positive values indicate that the participant uses the ergative marker more frequently than average (indicated by the vertical line at x = 0). For instance, participant 1r uses the ergative marker most often, whereas participant 1t has used it less often than the "average" participant. We have additionally added Year as a colour. This shows that there is no significant increase in the use of the ergative marker over the year. Overall, Figure 4 reveals that there is a great deal of variation among the different learners.

Learners from the same year can show a very different marking pattern. This finding is further analysed in the qualitative part of our study.

Of course, frequency of use does not necessarily imply that the uses are also correct. To evaluate this, we additionally examined the proportion of correct uses for each learner. The results of this analysis are given in Table 6.

Table 6: Correct use of the ergative marker *-ne* for each learner (that produced the ergative marker in a perfective sentence). Correct also includes correct usages of a non-ergative marker. The order of the Participants follows that of Figure 4.

Participant (Year)	Correct % (Correct / Total)	Correct ergative / Total ergative
1r (3)	90% (52/58)	34/38
1p (2)	89% (41/46)	20/21
1n (2)	89% (16/18)	11/13
1v (3)	79% (19/24)	11/12
1w (4)	100% (39/39)	23/23
1s (3)	85% (22/26)	13/13
1q (3)	67% (38/48)	23/26
1x (4)	89% (24/27)	18/18
1y (4)	91% (62/68)	34/34
1u (3)	59% (22/37)	6/8
1o (2)	72% (18/25)	2/2
1t (3)	58% (29/50)	2/2

Overall, the learners use the ergative marker correctly. Learners 1r, 1p, and 1q are the most frequent users of the ergative marker. 4th-year students (1w, 1x, 1y) have fewer utterances but all ergative markers are used correctly. By correct use, we mean the use of *-ne* marking in perfective, transitive contexts, and null-marking in perfective, intransitive contexts.

Individual learner profiles

As we can see from the table above, learners from the same year can portray quite different *-ne* marking strategies. Based on the utterances of the learners, we have discerned 5 types of learner behaviour.

1. No perfective, no *-ne* marking (n = 3). Our quantitative analysis already confirmed that there was no use of the perfective verb form, nor of the ergative case marker *-ne* in the utterances of the learners of year 1 (learners 1k, 1l and 1m). When they produced utterances in the past tense, they used the past imperfective. [authors' remark: the number 84 refers to the utterance ID in the corpus data. The dataset will be shared after acceptance for publication in the data repository TROLLING]

- 84: laḍkī aur laḍkā aurat=ko *dekhtā + the

  girl and boy.NOM woman.OBL=OBJ see-IMPERF

  'The girl and the boy saw a/the woman.'

2. Perfective, preference for *zero* marking (n = 3). Learners 1o, 1t, and 1u have acquired the use of the perfective verb form. With regard to the subject of the transitive verb, the data show that these learners are to some level aware of the *-ne* marking rule as they do utter a few *-ne* marked subjects (correct uses as well as overgeneralisations). However, the learners do not actively master this knowledge as they mainly produce zero-marked subjects. It is noteworthy that the production of the *-ne* marked subjects correlates to either the interviewer's question "can you continue in the past tense" or to a comment by the participant realising he/she has been making mistakes with *-ne* marking.

- 431: ***yah** **ḍākū** *iskā ūṃṭ uṭhāyā

  This.NOM bandit.NOM his.NOM camel.NOM lift-PERF

  'This bandit lifted his camel'

3. Acquisition of *-ne*: omissions and overgeneralisations (n = 5). The data of learners 1p, 1q, 1r, 1s, and 1x suggest that these learners are aware that there is an opposition between the marking of the subjects of transitive perfective verbs and the subjects of intransitive perfective verbs. There is also a clear association of *-ne* marking with perfective verb forms. However, this opposition is not stable and the learner data portray fluid language production with omissions as well as overgeneralisations, to varying degrees. It is noteworthy that all of the learners produce errors of omission but not all the learners produce overgeneralisations. Some of the learners (learners 1s and 1x) that omit the *-ne* marker rather than overgeneralising its use, also portray a preference for using the imperfective over the perfective verb form. Learner 1q produced all error types, whereas learner 1r mainly produced overgeneralisations to intransitive perfectives.

- 670: kyoṃki *āp merī madad kiyā

  Because you.NOM my.NOM help.NOM do.PERF

  'Because you helped me.'

- 737: us=ne gir + paḍ-ā

He.OBL=ERG fall + AUX-PERF

'So she fell.'

4. Drop of subject with perfectives (n = 2). Both learner 1n and learner 1v portray knowledge of *-ne* marking but like the learners above, regularly omit or overgeneralise the marker. Additionally, these two learners drop the subject with perfective verbs (as opposed to dropping subjects of imperfective verbs). Although they are certainly not the only learners in our sample to do so, these two learners stand out because they drop the subject equally as much as they realise overt subjects with perfective verbs:

- 1471: uskī madad *kiyā

  her.NOM help.NOM do-PERF.SG

  '(He) helped her.'

5. Acquisition of *-ne:* target-like proficiency. The data of learners 1w and 1y portrays a clear opposition between subjects of transitive perfectives and intransitive perfectives (cf. 1510, 1605). Learner 1w is the first learner of the sample who neither omits nor overgeneralizes *-ne* marking. We can thus assume that this learner has acquired *-ne* marking. Learner 1y, on the other hand, still omits the ergative *-ne* with transitive perfectives on a few occasions, but given the high number of perfectives, and the high number of correctly *-ne* marked subjects, we do not find these instances to indicate that the learner has not acquired *-ne* marking.

- 1510: to cūh-ā aur ek ādmī bhāgne + lage

  So mouse-NOM and one man.NOM flee.INCH

  'So the mouse and one man started to flee.'

- 1605: to tum=ne ek liphāph-ā dekh-ā

  So you=ERG one envelope-NOM.SG see-PERF.SG

  'So you saw a letter'

To summarize, no ergative *-ne* marker was produced by first year learners. The *-ne* marker was used only 4 times in non-perfective verb contexts. The frequency of perfective verb contexts increases throughout year 2, 3 and 4, with the highest amount of overgeneralisations to intransitive perfective verb contexts in year 3. When analysing the data from a qualitative angle, we can determine 5 learner profiles:

- Profile 1: No perfective, no *-ne* marking (n = 3);
- Profile 2: Perfective, preference for *zero* marking (n = 3);
- Profile 3: Acquisition of *-ne*: omissions and overgeneralisations (n = 5);
- Profile 4: Drop of subject with perfectives (n = 2);
- Profile 5: Acquisition of *-ne:* target-like proficiency (n = 2).

6. Results: -ko marking

Our participants produced 696 (38%) utterances that contained a patientive argument, 109 (6%) of which were also used with the differential object marker *-ko*. Figure 5 shows the usage of the DOM by year:

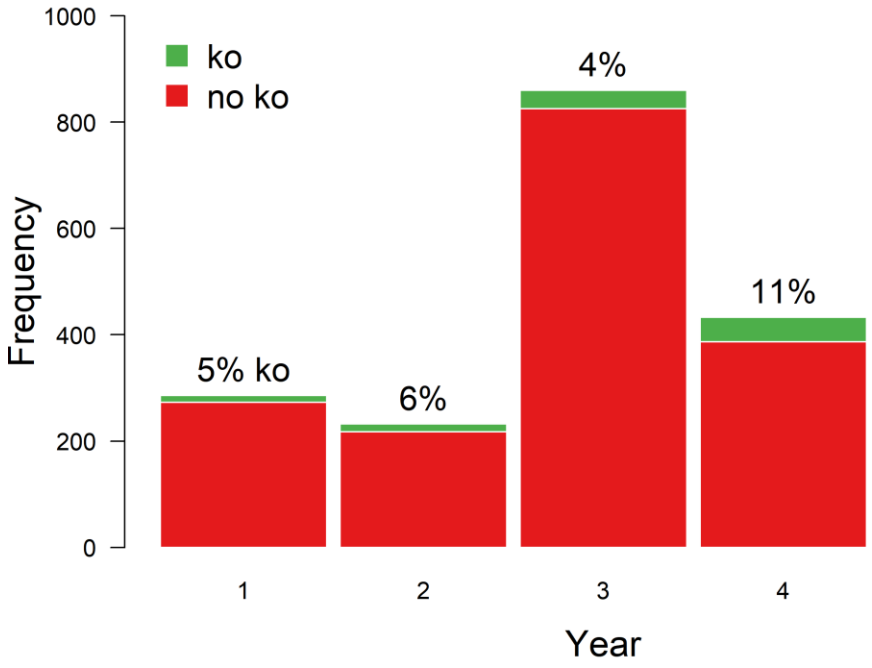

Figure 5: Usage of *-ko* by Year. Bar heights indicate absolute frequencies. Green refers to *-ko* uses, red to non *-ko* uses. The proportion of *-ko* markers conditional on Year is added as a percentage.

Table 7 shows that the DO marker *-ko* is never used with a non-specific DO. Moreover, the table shows that the majority of the specific DOs remain unmarked, although a gradual increase in *-ko* marked DOs is noticeable over the years.

Table 7: Usage of *-ko* vs. DO specificity by Year.

		DO marker	
	DO specific	**Yes**	**No**
Year 1	Specific	13	43
	Non-specific	0	29
Year 2	Specific	15	44
	Non-specific	0	42
Year 3	Specific	35	130
	Non-specific	0	166
Year 4	Specific	46	53

Non-specific	0	80

Limiting the data to sentences with a specific DO, Table 8 and Figure 6 show the relation between the usage of *-ko* and DO animacy. Recall that *-ko* should be used in all animate contexts.

Table 8: Usage of *-ko* in relation to DO Animacy by Year. The data is limited to sentences with a specific DO.

		Ko as DOM	
	DO Animacy	**Yes**	**No**
Year 1	Animate (human)	6	5
	Animate (non-human)	1	13
	Inanimate	6	25
Year 2	Animate (human)	10	7
	Animate (non-human)	4	11
	Inanimate	1	26
Year 3	Animate (human)	24	11
	Animate (non-human)	9	28
	Inanimate	2	91
Year 4	Animate (human)	18	2
	Animate (non-human)	8	6
	Inanimate	20	45

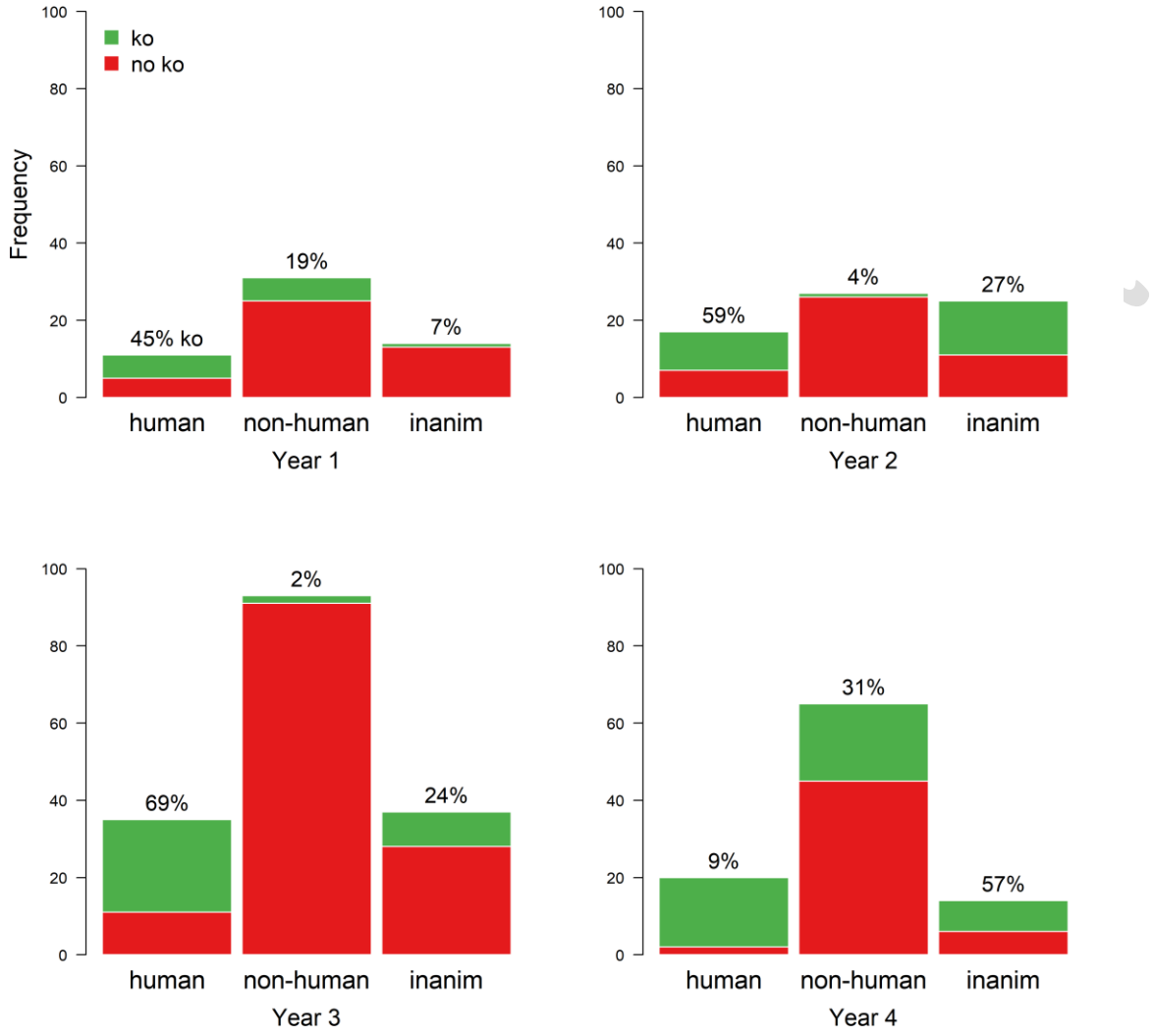

Figure 6: Use of *-ko* marker in relation to Animacy by Year. Bar heights indicate absolute frequencies. Green refers to *-ko* uses, red to non *-ko* uses. The proportion of *-ko* markers conditional on Animacy by Year is added as a percentage. The data is limited to sentences with a specific DO.

We fit a mixed-effects logistic regression model with the usage of the DO marker *-ko* as the outcome variable, an interaction effect between DO animacy and Year as a fixed factor, and with Learner as a Random effect. The interaction is significant at the 5% significance level based on a likelihood ratio test of fitted models (G = 15, df = 6, P-value = 0.01), which suggests that the effect of DO animacy on the

usage of *-ko* is not the same for each year. Figure 7 visualizes the interaction between DO animacy and Year:

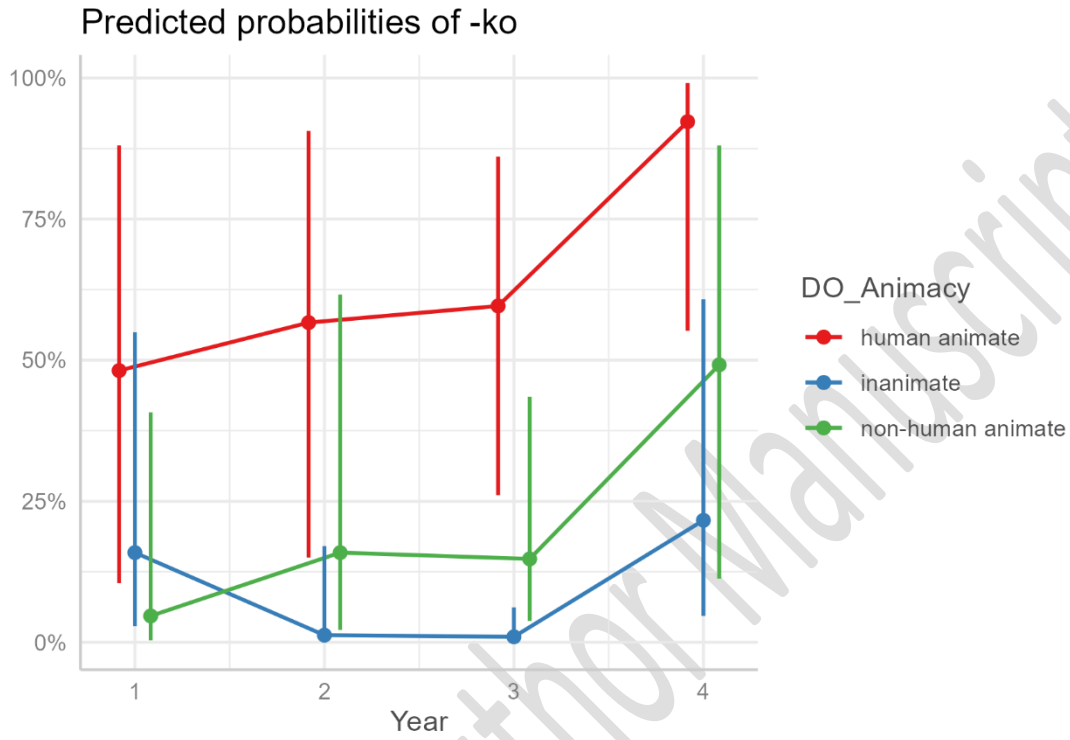

Figure 7: Effect plot for the interaction effect between DO Animacy and Year on the usage of the DO marker *-ko*.

Throughout all years, the DO marker *-ko* is mostly used with a human animate DO, except in year 1, where inanimate and non-human animates have an opposite effect. It is the latter which results in a significant interaction, as the lines between years 2, 3 and 4 are nearly parallel. Second and 3rd-year learners show similar uses, and both have a preference for *-ko* marking with a (specific) human animate DO (recall that we are looking at the subset with specific DOs). The results further indicate that although fourth-year learners use the DO marker predominantly with human animate DOs, there is also an increase with inanimate and non-human animate DOs.

Table 9 offers a posthoc comparison of the difference between the means associated with each level of Animacy, controlling for year:

Table 6: Posthoc pairwise differences for each level of Animacy, conditional on Year. The estimates are on the probability scale.

Year	Contrast	Est. (SE)	z ratio	P-Value

1	human anim – inanim	32% (20%)	1.598	0.25
	human animate – non-human anim	43% (24%)	1.805	0.17
	inanim – non-human anim	11% (11%)	1.045	0.55
2	human anim – inanim	55% (24%)	2.276	0.06
	human animate – non-human anim	41% (20%)	2.051	0.10
	inanim – non-human anim	-15% (14%)	-1.063	0.53
3	human anim – inanim	59% (17%)	3.426	0.002
	human animate – non-human anim	45% (13%)	3.373	0.002
	inanim – non-human anim	-14% (9%)	-1.519	0.28
4	human anim – inanim	71% (12%)	5.958	< 0.0001
	human animate – non-human anim	43% (23%)	1.905	0.13
	inanim – non-human anim	-28% (18%)	-1.508	0.28

The posthoc comparison shows that for learners of year 3 and 4, the *-ko* marker is significantly associated with an animate DO (in year 3, this also includes human animate DOs). The effect is strongest in year 4: human animate DOs have a 71% higher probability of being marked with *-ko* than an inanimate DO.

The individual learner effect is visualized in Figure 8, which shows that the DO marker *-ko* is most frequently used by learners 1y, 1p, 1v, 1t, and 1k.

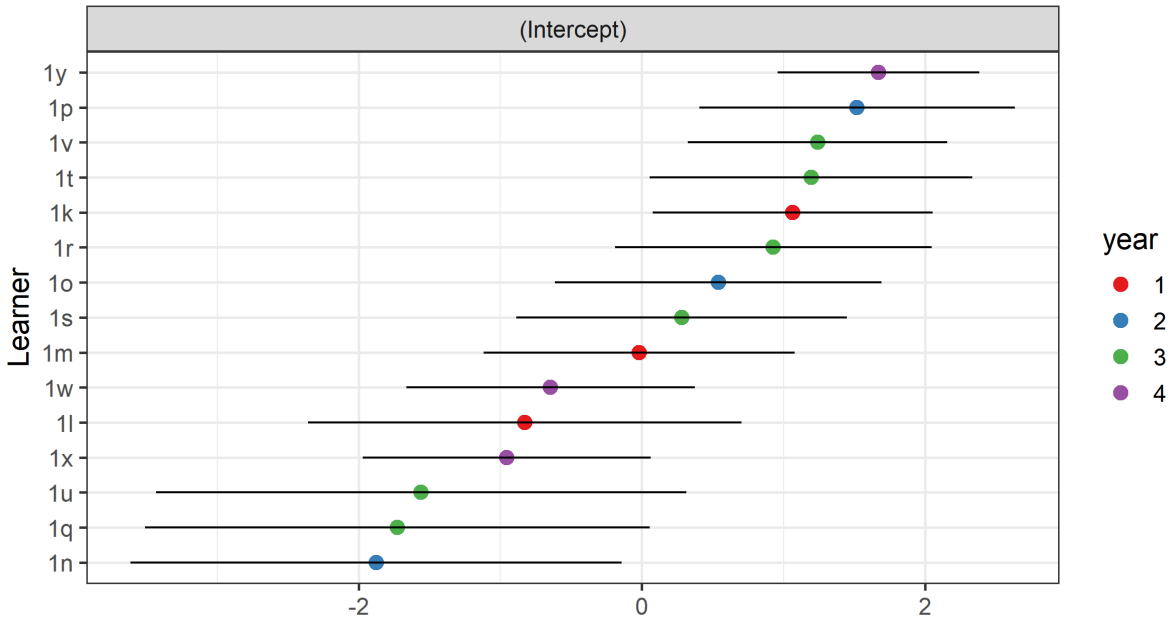

Figure 8: Best Unbiased Linear Predictors (BLUPS) for the random effect Learner.

Similar to the results of ergative marking, we notice that there is a high level of variation among the learners, with learners who never use *-ko* marking to learners who regularly use this marker. Again, we see that within the different years, different types of learner behaviour occur. Interestingly, when we compare the DO *-ko* to other uses of this marker, plotted for all the learners together, we immediately see that the use of *-ko* for the experiencer and the recipient is quite high. When we plot the data per year, we notice that here too, there is an increase in the usage of *-ko*, cf. Figure 9:

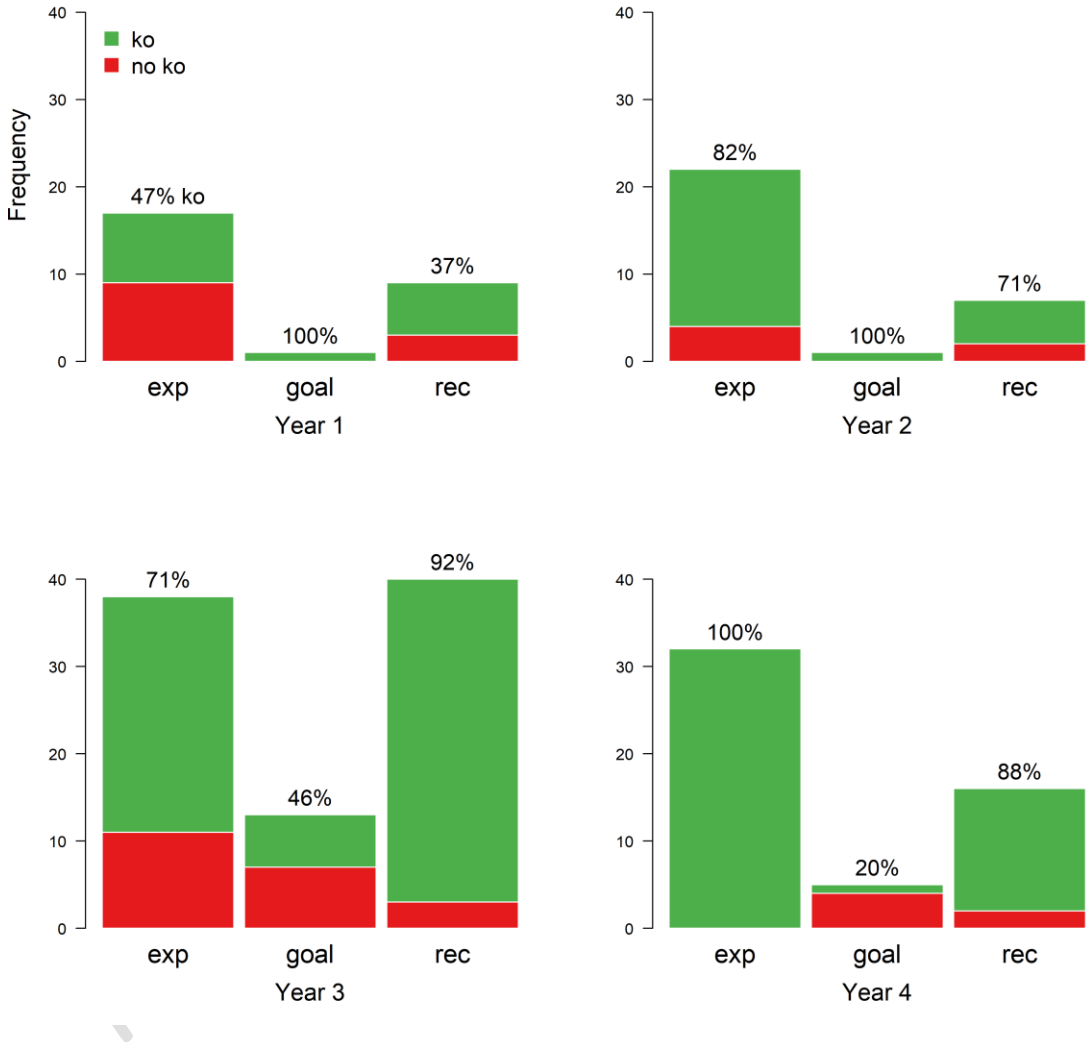

Figure 9: Other uses of *-ko* marker in relation to Semantic role by Year. Bar heights indicate absolute frequencies. Green refers to *-ko* uses, red to non *-ko* uses. The proportion of *-ko* markers conditional on Semantic role by Year is added as percentage.

Both Year and Semantic Role are highly significant main effects in a logistic regression model (see Appendix for model details). Both effects are visualized in Figure 10:

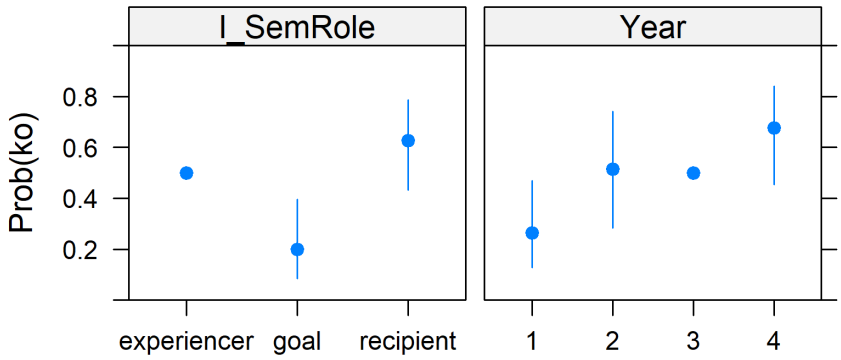

Figure 10: Predicted probabilities for Year and Semantic Role

The use of *-ko* is mostly associated with the role of experiencer and recipient, and less so with goal. It is important to note that there were very few utterances with goal objects overall. We also see an increase in the use of *-ko* from year 1 to year 4, with a plateau at year 2 and 3. An interaction effect between Year and Semantic Role was not significant, which means that we have no evidence that the effect of Semantic Role changes over the years.

Individual learner profiles

Even though we attested an increase in *-ko* marking throughout the different years of study, the participants' use of this marker remained quite restricted when it comes to marking of the DO. Based on a qualitative analysis of the data of each learner, we were able to discern 5 types of learner behaviour under which we could group the different learners. It immediately becomes clear that year of study does not necessarily warrant one type of learner behaviour.

1. No use of *-ko* as a DO marker (n = 2). *-Ko* as a DO marker has not emerged in the interlanguage of learner 1n and learner 1u. Interestingly, learner 1n does use *-ko* to mark the recipient and the experiencer construction in several utterances, although the marker is also dropped with these roles. The same applies to learner 1u, who uses *-ko* to mark experiencers, recipients, and goal objects but also regularly drops the marker (cf. 1336):

- 1336: laḍk-ā *laḍk-ī ***kutt-ā** detā + hai

  Boy-NOM girl-NOM dog-NOM give-IMPERF + AUX

  'The boy gives the dog to the girl'

2. Emergence of *-ko* DO-marking (n = 4). Learners 1k, 1l, 1m, and 1q all use *-ko* as a DO marker, be it in varying ways. Learners 1l and 1q, e.g., use the marker once, whereas learners 1k and 1m use the marker multiple times. Strictly speaking, learner 1q does not even use the marker as she uses a syncretised version of the pronoun *usko,* i.e. *use*:

- 635: use mat chūn-o

  He.OBJ Not touch-IMP

  'Don't touch him.'

We decided to include her here as she also produces utterances such as the following, where the DO is in the oblique case (which requires the presence of a case marker):

- 736: Aaricia=ne is patthar nahīṃ dekh-ā

  Aaricia=ERG this.OBL stone.INV no see-PERF.NOM.SG

  'Aaricia didn't see this stone'

In all of the *-ko* marked utterances of learner 1k, 1l, and 1m, the DO is either a human animate patient (usually the patient of a highly transitive verb such as *cūmnā* or *mārnā*) or the *-ko* marked DO is preceded by *-ko* marking of another semantic role, such as the recipient or the experiencer construction (86-87).

- 86: aurat laḍk-*ā=**ko** patr *dātā + thā

Woman.NOM boy-*NOM=OBJ letter give-IMPERF + AUX

'The woman gave the letter to the boy'

- 87: aurat laḍk-*ā=**ko** cūmtī + thī

  Woman.NOM boy-NOM=OBJ kiss-IMPERf + AUX

  'The woman kissed the boy'

3. *-ko* marking: grouping of non-human animates and inanimates (n = 7). Learners 1o, 1p, 1r, 1s, 1t, 1w and 1x all show similar behaviour with regard to *-ko* marking of the DO. They all associate *-ko* marking with the human animate, specific DO (962). Interestingly, for all of these learners[8], we attested a correlation between the use of the numeral *ek* and zero-marking. With some learners, all of the unmarked human animate DOs are realised with the numeral *ek* (and thus annotated as non-specific, cf. infra) (1039). Additionally, they seem to group non-human DOs and inanimate DOs, both of which only get marked in a few instances and only when they are specific. Note that in the data of these learners, only 1 or 2 specific non-human animate/inanimate DOs are being marked, whereas *-ko* marking is almost always present with specific human animate DOs. As such, the data of these learners show two oppositions: 1) between human animates on the one hand, and non-human animates and inanimates on the other, and 2) between specific and non-specific DOs.

- 962: to Tim Paulus=ko dekh + rahā + thā

  So Tim.NOM Paulus=OBJ see + PROG + AUX

  'So Tim was looking at Paulus'

- 1039: maiṃ ek ronevālī laḍkī dekhtī + hūṃ

  I.NOM one cry.NON-FIN girl.NOM see-IMPERF + AUX

  'I see a weeping girl'

As for learners 1t and 1x, we also attested that they produced quite some pronominal DOs (such as the proximate *yah* and the obviate *vah*). Interestingly, they marked the animate pronominal DOs with *-ko*, but left the inanimate pronominal DOs unmarked.

- 1200: vah puppy to hāth=meṃ to
  She.NOM puppy.NOM so hand=LOC so
  is=ko rakh + detā + hai
  this.OBL=OBJ place + AUX
  'So the puppy, she places it in (her) hands.'

- 1679: marusthal=kī adivāsī=ne **yah** nahīṃ
  desert=GEN nomad=ERG this.NOM no
  samajhā
  understand-PERF.NOM
  'And the desert nomad didn't understand this'

4. *-ko* marking: opposition between animates and inanimates (n = 1). Learner 1v shows a different type of behaviour for *-ko* marking of the DO. In the data of this learner, an opposition seems to have arisen between animate and inanimate DOs. Again, this does not mean that all animate DOs are being marked (omissions do occur) but there is a clear difference in the frequency of *-ko* marked DOs. Whereas almost all the human animate DOs were marked, and half of the non-human animate DOs (cf. 1399), only 1 inanimate DO was marked with *-ko*. Here too, an association with specificity seems to have been established because for all of the DO types only specific DOs were marked with *-ko*.

- 1399: islie cār ādmi-yoṃ=ne ūṃṭ=ko lāye
  because four man-OBL.PL=ERG camel=OBJ bring-PERF.PL
  oase=par
  oase=LOC
  'Because of this, the four men took the camel to the oasis.'

5. *-ko* marking: decrease in omission of *-ko* for ALL of the roles (n = 1). Learner 1y is the first and only learner of this study that uses *-ko* to mark almost all of the animate DOs, and regularly marks inanimate specific DOs:

- 1907: aur laḍk-e=ne **us** **phūl=ko** laḍkī=kī
  And boy-OBL=ERG that.OBL flower=OBJ girl=GEN.F
  bāl-oṃ=meṃ rakh-ā
  hair-OBL.PL=LOC put-PERF.M.SG
  'And the boy put that flower in the girl's hair.'

The only unmarked human animate DO *ādmī* is realised with the numeral *ek*. However, the learner regularly produces utterances with a *-ko* marked DO that are realised with *ek*[9] (even with the same lemma, cf. 1855), both animate and inanimate. Moreover, when we have a look at the unmarked DOs, there are quite some DOs that are realized with the numeral *ek* but more than often they are realised without the numeral. Thus, we cannot assume an association between the numeral *ek* and zero-marking for this learner. Together with learner 1x, this learner is the only learner from this study who marks both the DO and the recipient with *-ko* in the same utterance. Notice how the subject is not marked with the required ergative case marker *-ne*:

- 1855: maiṃ **ek** **ādmī=ko** dekhtā+hūṃ
  I.NOM one man=OBJ see-IMPERF+AUX
  'I see one man.'

- 1894: aur laḍkā **us** **kutt-ā=ko** **laḍkī=ko**
  And boy.NOM.SG that.OBL.SG dog-NOM=OBJ girl=OBJ
  diyā
  give-PERF.M.SG
  'And the boy gave that dog to the girl.'

To summarize, throughout the data, there were no instances of DO *-ko* marking with a non-specific DO. Overall, the data shows high omission rates for *-ko*, even when the DOs are (human) animate and specific. Interestingly, DO *-ko* emerges already in the first year. Based on an individual learner analysis, we distinguished the following learner profiles:

- Profile 1: No use of *-ko* as a DO marker (n = 2);
- Profile 2: Emergence of *-ko* DO-marking (n = 4);
- Profile 3: *-ko* marking: grouping of non-human animates and inanimates (n = 7);
- Profile 4: *-ko* marking: opposition between animates and inanimates (n = 1);
- Profile 5: *-ko* marking: decrease in omission of *-ko* for ALL of the roles (n = 1).

Additionally, we mined the data for uses of non-DO *-ko*, and, although the data was rather limited, found that there was a high rate of (accurate) marking of the recipient and the experiencer subject.

7. Discussion

This study investigated the development of *-ne* marking and *-ko* marking with Dutch-speaking HFL learners. On a descriptive level, our study aimed to examine how L1 speakers of Dutch, a typologically distant language which lacks both ergativity and DOM, acquire Hindi split ergativity and DOM. On a theoretical level, our study aimed to answer the question if processing constraints can predict the emergence of a linguistic structure and whether the trajectory follows developmental stages, or if variation is too high to discern such patterns. We follow Hulstijn (2015) and included accuracy measures to account for the intra-stage developments, and describe what happens once the linguistic structure has emerged.

Group results -ne marking and -ko marking

By means of an analysis of the developmental trajectory of the learners and inter-learner variation, we established the effect of semantic and syntactic constraints on the learning of these case markers. The results show that general patterns apply when we look at the learner data per year of study, i.e. we attested an increase in the use of each case marker over the different years of study, and this incremental process followed a more or less similar route, which may be summarized as follows.

***-ne* marking**: The data showed that the use of the ergative case marker *-ne* was first associated with perfectivity, whereas the opposition between transitivity and intransitivity was acquired later. The data also showed that the level of correctness gradually increased towards a high level of proficiency in year 4. The observed dynamics with regard to *-ne* marking bears resemblance to several earlier findings in studies of HFL. Baten & Verbeke (2015) performed a cross-sectional study with Dutch-speaking HFL learners who had been learning Hindi for one, two, three, or four (or more) years. They found that the

learners pass through three stages: no use of *-ne*, overgeneralization of *-ne* (mainly to intransitive perfectives), and finally acquisition of *-ne*. According to their study, the learners associate *-ne* marking with perfectivity rather than transitivity. Similar findings come from a cross-sectional study by Ranjan (2016), who looked at the acquisition of *-ne* with English-speaking learners, which he divided in an intermediate and an advanced group. His results similarly suggest that the learners acquired the perfective aspect first (marking both transitive and intransitive subjects) and that the consistent use of *-ne* gradually increases with more advanced learners of Hindi.

***-ko* marking**: With regard to *-ko* marking, as non-specific DOs are never marked with *-ko* by the learners in our study, specificity seems to be the determining feature for the emergence of *-ko* in the IL of HFL learners. However, specificity clearly interacts with animacy. Learners first appear to associate *-ko* marking of the DO with human animate specific patients. Gradually, an association with non-human animate specific patients and inanimate specific patients emerges. First non-human animate specific patients are treated in a similar way as inanimate patients, i.e. they remain null-marked. Subsequently an opposition between animate specific patients and inanimate specific patients appears to arise, and as a result, an increase of *-ko* marked arguments in general and hence a much clearer opposition between the *-ko* marking of specific and the null-marking of non-specific patients. Here, too, the frequency of *-ko* marking of DOs increases towards year 4, although the omission rate stays much higher when we compare the results to the results of *-ne* marking. The findings on DOM here are in line with Ponnet et al. (2016) who focused on the acquisition of DOM with HFL learners from different language backgrounds and in a study abroad context. The learners had different amounts of exposure, having studied Hindi for one, three, five, or seven semesters. In general, Ponnet et al. (2016) found that learners tend to omit *-ko* marking and that the use of the marker gradually expands from animate/specific DOs to inanimate/specific DOs. They also found that the learners' strategy differ: some learners focus more on the feature animacy, whereas others associated *-ko* marking with specificity. For *-ko* marking of other roles than DO, we notice that all of the learners use *-ko* to mark recipients and experiencer constructions (whereas for *-ne* marking and *-ko* marking of the DO, several learners did not use these markers). Here, too, we see a gradual increase towards a target-like usage of *-ko* marking.

Individual learner results -ne marking and -ko marking

Looking beyond the general patterns, we followed the notion that each learner has their unique interlanguage by focussing on the individual learners. Our multiple case study analysis revealed that patterns of marking may vary per speaker but also according to which particular marker we are focusing on.

***-ne* marking**: As for *-ne* marking, we found that within the individual learners' samples, overgeneralisations to imperfective verbs do occur, but are rare. Errors within perfective verb constructions are much more frequent, i.e. no *-ne* marking with transitive perfectives, and to a lesser extent, *-ne* marking with intransitive perfectives. This implies that the learners easily associate ergative case marking with perfectivity, but that the syntactic/semantic restrictions within (transitive vs. intransitive) take longer to be acquired. Moreover, before the learners start using *-ne*, they first acquire the perfective verb form, as several learners used the perfective verb form but did not use *-ne*. Additionally, some learners tended to associate the use of perfective verb forms with subject-dropping.

***-ko* marking**: Concerning *-ko* marking, the data showed that the learners struggle to apply the different semantic and syntactic constraints that guide differential case marking in Hindi. We noticed inter-learner variability and variability with regard to the different roles that are marked by *-ko*:

1. Each of the learners of year 1 had some notion of the use of *-ko* as a DO marker. However, when we look at the individual learners of year 2 and even year 3, some learners either did not use this marker at all or did only on a few occasions. Each time, the *-ko* marked arguments involved human animate specific patients.

2. All learners used *-ko* as a marker of recipient and experiencer constructions. The learners who did not use *-ko* as a DO marker, did use *-ko* for other semantic roles. This could imply that, for these learners, *-ko* is associated first with the recipient/experiencer constructions and only later extended to a DO marker.

3. Non-specific DOs in our dataset were not marked with *-ko*. This implies that, for Dutch-speaking HFL learners, the emergence of *-ko* marking is first associated with specificity. It is however clear that specificity interacts with animacy, because *-ko* emerges first with human animate specific DOs, and it is never the case that all specific DOs are being marked. We therefore assume that an association with specificity is established first, but that within the category of specific DOs, animacy plays an important role. More specifically human animacy, which then expands to non-human animate arguments. This transition goes hand in hand with the opposition between specific and non-specific arguments, as more types of specific DOs are being marked.

4. In the dataset, goal objects expressing direction were scarce (20 observations), but, interestingly, all of the animate goal objects were marked with *-ko*. Quite unsurprisingly, all of the recipients and experiencer nouns were animate nouns. The actions expressed with these semantic roles are usually associated with animate nouns. This might trigger an animacy bias for *-ko* marking, also in DO-context.

All in all, once each marker has emerged, and even though errors occur, which show that the learners struggle with the different semantic and syntactic constraints, the learners know that their use of the case markers is affected by these different constraints. It is the combination of the different constraints that poses challenges, especially in a context where the learners are focused on telling the stories. It is interesting to notice that in some of the elicited data there are certain cognitive triggers where the learners realise they have been making mistakes or should be using different grammatical constructions. This was especially clear for *-ne* marking. However, in a semi-spontaneous experimental setup like the one we have chosen, it is not possible to know what exactly triggers this (meta-linguistic) knowledge.

Omission of -ne and -ko marking

The question remains why the omission rates remain high for several of the advanced learners, in particular regarding DOM. In their study with Hindi heritage speakers on case marking and agreement, Montrul et al. (2012) found that heritage speakers have difficulties with *-ne* marking, DOM, and OV-agreement. Interestingly, omission rates were higher for *-ne* marking, which is different to what we found in this study. In a subsequent study, Montrul et al. (2019) compare Hindi heritage speakers with HFL learners. Both speaker groups had similar proficiency rates. Here, too, heritage speakers tended to omit the *-ne* marker, whereas the HFL learners were more prone to overgeneralisations. For the

different types of *-ko* marking, both learner groups showed similar results which meant that both groups omitted the marker in contexts where its use was required. HFL learners are instructed learners, which in the case of Montrul et al.'s (2019) study, could account for the difference between the two speaker groups – moreover, the results of the HFL learners of Montrul et al.'s study align with what we found in this study. In both the study with heritage learners and the study comparing early (heritage speakers) and late bilinguals (by which they mean foreign language learners) and balanced bilinguals, Montrul et al. (2012; 2019) focus on accuracy levels and measure these against the language of native speakers. They therefore do not include a detailed analysis of the instances where the learners overgeneralize or omit the markers, and did not mention individual differences between the learners. In future research, it would be interesting to set up an experiment that compares native speakers of Hindi with non-native speakers from different backgrounds and different learning environments, which examines the different factors that might interact with perfectivity/transitivity and animacy/specificity in the language of these different groups of speakers.

With regard to the possible influence of the L1 of the learners, we are careful in our conclusions. Earlier work on the role of transfer in languages that have differential case marking, such as studies on the acquisition of L2 Spanish and L2 Romanian (Montrul 2019b; Montrul & Gürel 2015) have stressed the advantage that speakers of DOM-languages have when acquiring another DOM-language. Earlier studies on the acquisition of Hindi (which we discuss in detail below) have shown that English-speaking learners and Dutch-speaking learners of Hindi show high omission rates for *-ne* marking and *-ko* marking, and state that this may be due to transfer of native language structures (English and Dutch do not portray overt case marking). To determine if transfer is at play, we need to be able to contrast different learner groups with a highly controlled set of L1-L2 combinations (which leaves scope for an exciting area of research in the future). Moreover, in an earlier study by Ponnet et al. (2016), a group of learners that had Dutch as an L1, and a group of learners from different language backgrounds were presented the same spontaneous oral production task, and it appeared that there was no significant difference between the two learner groups with regard to Hindi DOM. Instead, Ponnet et al. found language internal explanations for the marking behaviour of their participants, which is similar to the findings of our current study.

Developmental stages or variation?

How do the findings of our study fit into the theoretical notions of developmental stages and variation? The PT framework has a strong focus on the individual learner and predicts developmental stages via feature unification and mapping of forms onto functions or positions (Pienemann, 1998). To assess if a learner has entered a certain stage, PT adopts the emergence criterion (Pallotti, 2007). As such, a high rate of accuracy for a certain linguistic feature is not required: the moment the feature has emerged, the learner is considered developmentally ready to acquire the forms that belong to that stage. Traditionally, PT studies place the cut off point for emergence, i.e. how many times a certain form should occur for it to be considered emerged, at three contexts where the form is used. In section 3 of this article, we explained in detail the route for trajectory predicted by Baten & Ponnet (2023) that we follow in this study, represented here in table 10:

Table 10: developmental trajectory

Processing procedures	**Case in Hindi**
4. S-procedure	Functional nominative marking

	-ne marking agents of transitive perfectives *-ko* marking of animate DO *-ko* marking of non-human animate DO
3. Phrasal procedure	*-ko* marking specific DO
2. Category procedure	*-ne* marking of agentive-like arguments *-ko* marking of recipient *-ko* marking of experiencer
1. Word/lemma access	Default null-marking

This trajectory is based on the processing constraints that L2 learners encounter when learning a language, and based on Levelt's (1989) model of language generation and LFG (Bresnan et al. 2015) that both assume a lexically driven language processor. Functional elements of the interlanguage only gradually become available throughout the different stages as described in table 10, making specific types of information exchange available only at different stages of the IL development.

Does our study verify or falsify the predictions made by Baten & Ponnet (2023)? The analysis of the individual learner data revealed that case markers emerged gradually and were increasingly mapped according to certain features. In their study, Baten & Ponnet (2023) predict that *-ne* marking will initially be mapped onto agentive-like arguments. This prediction is borne out by the data of our study, with the important qualification that agents of imperfective verbs barely received *-ne* marking. We therefore surmise that *-ne* marking is initially mapped onto agentive-like, perfective arguments. Moreover, given that some learners produced perfective verb forms while they did not produce any *-ne* marked argument, we assume that perfectivity acts as a 'developmental trailer' (Pienemann, 1998, p. 250) for the mapping of *-ne* onto agentive arguments. The data also confirmed an early emergence of *-ko* marking for recipients and experiencer subjects. Baten & Ponnet (2023) predict that with regard to DOM, *-ko* will first be linked to specificity. Here too, our data confirms this prediction in two ways. Firstly, in our data, we saw that, initially, *-ko* emerges with human animate specific direct objects only. Secondly, there is an early association of zero-marking with non-specific arguments, especially as non-specific arguments never receive *-ko* marking in our data. It appears as if the zero-marking of non-specific inanimate DOs, and the optionality of marking with specific inanimate DOS, render the extension of zero-marking to animate DOs. Only to be countered by a functional association of *-ko* marking when the feature specificity and (human) animacy coincide, resulting in a series of oppositions.

It is important to distinguish between emergence and mastery. Our individual learner profiles showed that, with the learners in this study, *-ko* marking emerges before *-ne* marking. However, by year 4, we found higher accuracy rates for *-ne* marking than for *-ko* marking. This brings us to the variation that we noticed while studying the HFL learners as multiple cases. The data of our study shows a gradual increase of accuracy for case marking with the learners of this study. With regard to the emergence of the Hindi case marking system, our study shows that assumptions on the basis of PT are able to make useful predictions about the route of learner development. However, language development doesn't limit itself to emergence. There is a long way between emergence and mastery, which is currently being explored in the PT literature by adding variationist approaches to the framework (Dyson, 2021; Pienemann, 2021, among others). Our analysis also brought to light varying rates of accuracy between

the different learners, as well as different types of learner behaviour behind the accuracy measures. Despite these varying rates of accuracy, we managed to link their learner behaviour to the different developmental stages as described by PT. Hence, developmental stages and between-learner variation are, if we adopt the PT viewpoint of emergence, not mutually exclusive. The idea of variation and systematicity in language acquisition is of course not new (Ellis, 2008). Applied to this particular set of learners, this two-way approach enabled us to investigate HFL in a more detailed manner as has hitherto been done in studies on HFL.

We have summarized the result of our analysis in table 11 below. We linked the different types of learner behaviour that we found in this study to the different PT-stages, which are illustrated by Table 11. The predictions by Baten & Ponnet (2023) have a strong focus on the role of the semantic and syntactic constraints on the order of emergence of the case marking system in HFL. In our analysis in Table 11 we include the role of the optionality of the markers to account for the route of development from emergence to mastery. For *-ne* marking, we noticed that the obligatory contexts are in fact identified early on by the HFL learners, but given the high rate of omissions, initially seem to be regarded as optional contexts. Important to note is that *-ne* is in fact used with a wide variety of verbs (which excludes the option that *-ne* is associated with a certain set of verbs initially). For *-ko* marking, based on what we found in the data we have verified that *-ko* marking of the DO is triggered by *-ko* marking of other roles, and as such we have placed the emergence of DO *-ko* marking already at the category procedure (at the same level as *-ne* marking, in particular because *-ko* had emerged in the IL of the year 1 learners whereas *-ne* had not). We thus expand on Baten & Ponnet's (2023) predictions and suggest that the optionality is expanded through a series of oppositions at the phrasal procedure. For the recipient and the goal object, there is no functional opposition with zero-marking and hence no optionality. For the DO, Baten & Ponnet (2023) proposed an association of *-ko* with the feature specificity. This is verified by what we found in our data. Beyond emergence, we find an early preference for zero-marking which continues throughout the different stages and might suggest an early acquisition of non-specificity, which emerges with sporadic *-ko* marking of animate, specific direct objects. Instead, at the phrasal procedure we place two emerging oppositions where the first argument type will receive *-ko* marking: 1) human animate, specific DOs vs. non-human animate and inanimate specific DOs, and 2) human and non-human animate, specific DOs vs. inanimate specific DOs. Finally, at the S-procedure, we assume a decrease in omission for all of the roles and find inanimate, specific DOs that are also overtly marked with *-ko*.

Table 11: variation within developmental trajectory

Processing procedures	**Case in Hindi**
4. S-procedure	*-ko* marking: decrease in omission of *-ko* for ALL of the roles Acquisition of *-ne*: target-like proficiency
3. Phrasal procedure	*-ko* marking: opposition between animates and inanimates *-ko* marking: grouping of non-human animates and inanimates
2. Category procedure	Recipient *-ko* marking Perfective, preference for *zero* marking Acquisition of *-ne*: omissions and overgeneralisations Drop of subject with perfectives Emergence of *-ko* DO-marking

1. Word/lemma access	No use of *-ko* as a DO marker No perfective, no *-ne* marking

8. Conclusion

In this study we looked at the acquisition of the Hindi case marking system by Dutch foreign language learners. We found that the learners acquire the markers through a process in which they first associate the use of the markers with one of the features, after which they usually extend its use to other features. In general, we see a gradual increase in the use of the markers over the four years of study, and especially for *-ne* marking, we see an increase of accuracy as well. We also found that, even though general patterns apply when looking at the data over the 4 different groups, a different picture arises when we analyse the data of each learner individually. Then we see that learners may perform well on the use of one case marker but might not use the other case marker at all. Additionally, different learners exhibit different types of learner behaviour and strategies with regard to the use of these case markers.

To conclude this study, we point out some of its limitations and add some suggestions for future research. We were careful in our conclusions concerning transfer of zero-marking from the L1 Dutch to the HFL case marking system. Future research with speakers from different backgrounds that combines different experimental methods might be able to discern L1 influence from influence of TL internal linguistic factors (such as the ones discussed in our study). A more detailed, longitudinal study on *-ko* marking, with more data on marking of the recipient and different experiencer constructions, might shed light on how the different roles of *-ko* interact during the acquisition process. Something that we also noticed, but did not include as it would exceed the scope of this study, is that the acquisition of the perfective verb and *-ne* marking interacts with verbal agreement. E.g., in Figure 3 we saw a lower effect of transitivity in year 3, which may be due to a U-shaped development of split ergativity. In a future study on split ergativity in HFL, it would be highly interesting to investigate how the acquisition of split ergativity interacts with agreement. We noticed, e.g., that learners scored fairly well on agreement as long as utterances contained imperfective verb forms, or utterances with a masculine singular subject or DO, but that they have difficulties when the noun is feminine or plural. As both *-ne* marking and DO marking interact with agreement, it would be worthwhile to establish how the emergence and the development of these case markers interact with the development of the agreement system. Ranjan (2016) found that OV-agreement and default agreement emerge only after the *-ne* marker has been acquired. This raises questions that go far beyond the scope of the present study. Furthermore, longitudinal research is needed to verify if the processes observed in the current study also manifest themselves in real-time. For example, *-ne* marking had not yet emerged in the IL of the learners of year 1, whereas DO *-ko* marking had. However, the advanced learners in this study reached a much higher level of accuracy for *-ne* marking than for DO *-ko* marking. As Tarone (1988, p. 137) rightly points out, longitudinal studies are indispensable to “show the way in which variation at a single point in time is tied to longitudinal development of an interlanguage”.

Notes

[1] The South Asian languages Hindi and Urdu are closely related. As the students participating in this study were enrolled in a Hindi language course, at an institute where Urdu was not offered as a subject of study, we will mainly refer to Hindi in this article (as opposed to the term Hindi/Urdu that is used by many scholars).

[2] We do not use the term "absolutive" because Hindi does not conform to the ergative absolutive pattern (Butt & King 2004): in a split ergative language like Hindi both "nominative" and "absolutive" can designate the same form-function unit, the S-argument, and this leads to unnecessary confusion.

[3] Note that we will be using a consistent set of glosses even if examples are sourced from elsewhere. Our glosses may thus differ from the original source.

[4] Montrul et al. (2019), e.g., consider *-ko* marking to be obligatory when the DO is a personal pronoun, and optional when the DO is non-human and, in certain instances, human and non-specific.

[5] One of the limitations of this study is that it investigates so-called "Western Educated Industrialized Rich Democratic" (WEIRD) subjects (Henrich et al. 2010; Andringa & Godfroid 2020).

[6] Apart from learner 1m (year 1) who mentioned to have had some exposure to Nepali before enrolling in her Hindi course.

[7] See supplementary material for more details about the materials.

[8] Apart from learner 1o, who did not produce any non-specific human animate DOs at all.

[9] Learner 1y is not the only learner to do so, but he is the first learner who marks all three of the DO types that were realised with the numeral *ek*.

References

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

Andringa, Sible, & Godfroid, Aline. 2020. Sampling bias and the problem of generalizability in applied linguistics. *Annual Review of Applied Linguistics* 40. 134-142.

Baten, Kristof, & Saartje Verbeke. 2015. The acquisition of the ergative case in Hindi as a foreign language. In Kristof Baten, Aafke Buyl, Katja Lochtman, & Mieke Van Herreweghe (eds.), *Theoretical and methodological developments in Processability Theory*, 71-104. https://doi.org/10.1075/palart.4.04bat

Baten, Kristof & Ponnet, Aaricia. 2023. Extending PT to split ergative marking and differential object marking. In Satomi Kawaguchi, Bruno Di Biase and Yumiko Yamaguchi (eds.), *Processability and Language Acquisition in the Asia-Pacific Region* 9, 91-114. https://doi.org/10.1075/palart.9.04bat

Bates, Douglas, Mächler, Martin, Bolker, Ben, & Walker, Steve. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1). 1-48. https://doi.org/10.18637/jss.v067.i01.

Bossong, Georg. 1991. Differential object marking in Romance and beyond. *New analyses in Romance linguistics*. 143-170. https://doi.org/10.1075/cilt.69.14bos

Bresnan, Joan, Asudeh, Aash Asudeh, Toivonen, Ida & Wechsler, Steven. 2015. *Lexical-functional syntax.* John Wiley & Sons.

Butt, Miriam. 1993. Object specificity and agreement in Hindi/Urdu. *29th regional meeting of the Chicago Linguistic Society* 1. 80-103.

Butt, Miriam. 2006. *Theories of Case.* Cambridge: Cambridge University Press.

Butt, Miriam. 2017. Hindi/Urdu and Related Languages. In *The Oxford handbook of ergativity*. Oxford: Oxford University Press.

Butt, Miriam & King, Tracy Holloway. 2004. The status of case. In Veneeta Dayal and Anoop Mahajan (eds.), *Clause Structure in South Asian Languages*, 153-199. Springer: Dordrecht. https://doi.org/10.1007/978-1-4020-2719-2_6.

Caspi, Tal. 2010. *A dynamic perspective on second language development.* University Library Groningen dissertation.

Cenoz, Jasone & Gorter, Durk. 2019. Multilingualism, translanguaging, and minority languages in SLA. *The Modern Language Journal* 103. 130-135.

Dixon, Richard M. 1994. *Ergativity.* Cambridge: Cambridge University Press.

Douglas Fir Group. 2016. A transdisciplinary framework for SLA in a multilingual world. *The Modern Language Journal*. 100(S1). 19-47.

Dowty, David R. 1991. Thematic proto-roles and argument selection. *Language* 67. 547-619.

Dulieu, Jean. 1974. *Paulus de Boskabouter 7: Ali Baba en de Rover.* Bussum: Van Holkema & Warendorf.

Dyson, Bronwen Patricia. 2021. *Dynamic Variation in Second Language Acquisition: A language processing perspective.* Amsterdam: Benjamins.

Dyson, Bronwen Patricia & Håkansson, Gisela. 2017. *Understanding second language processing: A focus on Processability Theory.* Amsterdam: Benjamins.

Eberhard, David M., Simons, Gary F., & Fennig, Charles D. (eds.). 2021. *Ethnologue: Languages of the World.* (Fourth edition). Dallas, Texas: SIL International.

Evans, Nicholas, & Levinson, Steven C. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and brain sciences*. 32(5). 429-448. https://doi.org/10.1017/S0140525X0999094X.

Everaert, Christine. 2004. *Basisgrammatica van het Hindi.* Gent: Academia Press.

Fox, John. 2003. Effect Displays in R for Generalised Linear Models. *Journal of Statistical Software,* 8(1). 1-27. https://doi.org/10.18637/jss.v008.i15

García, Ofelia. 2011. *Bilingual education in the 21st century: A global perspective*. John Wiley & Sons.

Harrell, Frank E. J. 2021. rms: Regression Modeling Strategies. R package version 6.2-0. https://CRAN.R-project.org/package=rms.

Harrell, Frank E. J., & Dupont, Charles. 2021. Hmisc: Harrell Miscellaneous (4.5-0). https://CRAN.R-project.org/package=Hmisc.

Haznedar, Belma. 2006. Persistent problems with case morphology in L2 acquisition. *Interfaces in multilingualism*. 179-206. https://doi.org/10.1075/hsm.4.08haz.

Henrich, Joseph, Heine, Steven J. & Norenzayan, Ara. 2010. The weirdest people in the world?. *Behavioral and brain sciences*. 33(2-3). 61-83.

Heugh, Kathleen. 2013. Multilingual education policy in South Africa constrained by theoretical and historical disconnections. *Annual Review of Applied Linguistics* 33. 215-237. DOI:10.1017/S0267190513000135.

Hulstijn, Jan H. 2015. Discussion: How different can perspectives on L2 development be? *Language Learning* 65(1). 210-232.

Kachru, Yamuna. 2006. *Hindi* (Vol. 12). Amsterdam: Benjamins.

Klein, Udo & de Swart, Peter. 2011. Case and referential properties. *Lingua* 121. 3-19. https://doi.org/10.1016/j.lingua.2010.07.005

Larsen-Freeman, Diane. 2007. On the complementarity of chaos/complexity theory and dynamic systems theory in understanding the second language acquisition process. *Bilingualism: Language and cognition* 10(1). 35-37. https://doi.org/10.1017/S136672890600277X.

Levelt, Willem J. M. 1989. Speaking. From intention to articulation. The MIT Press.

López, Luis. 2012. *Indefinite objects: Scrambling, choice functions, and differential marking*. MIT press.

Malchukov, Andrej L. 2008. Animacy and asymmetries in differential case marking. *Lingua* 118. 203-221. https://doi.org/10.1016/j.lingua.2007.02.005.

Mardale, Alexandru. 2008. Microvariation within differential object marking. *Revue roumaine de linguistique* 53(4). 448-467.

Mardale, Alexandru & Montrul, Silvina (eds.). 2020. *The Acquisition of Differential Object Marking* (Vol. 26). Amsterdam: Benjamins.

McManus, Kevin. 2022. Methodology and design in corpus-based SLA research, European Second Language Association, Fribourg (Switzerland), 24 August 2022.

Mohanan, Tara. 1994. *Argument Structure in Hindi.* Stanford: Center for the Study of Language.

Montaut, Annie. 2004. *A grammar of Hindi.* München: Lincom Europa.

Montrul, Silvina A., Bhatt, Rakesh M. & Bhatia, Archna. 2012. Erosion of case and agreement in Hindi heritage speakers. *Linguistic Approaches to Bilingualism* 2(2). 141-176. https://doi.org/10.1075/lab.2.2.02mon.

Montrul, Silvina A. & Sánchez-Walker, Noelia. 2013. Differential object marking in child and adult Spanish heritage speakers. *Language Acquisition* 20(2). 109-132. https://doi.org/10.1080/10489223.2013.766741.

Montrul, Silvina A. & Gürel, Ayşe. 2015. The acquisition of Differential Object Marking in Spanish by Turkish speakers. In Tiffany Judy and Silvia S. Perpiñán (eds.), *The Acquisition of Spanish by speakers of less commonly studies languages*, 281-308. Amsterdam: Benjamins.

Montrul, Silvina A., Bhatia, Archna, Bhatt, Radesh M., & Puri, Vandan. 2019a. Case marking in Hindi as the weaker language. *Frontiers in psychology* 10. 461. https://doi.org/10.3389/fpsyg.2019.00461.

Montrul, Silvina A. 2019b. The acquisition of Differential Object Marking in Spanish by Romanian speakers. *Revista Española de Lingüística Aplicada* (RESLA). 32(1). 185-219.

Narasimhan, Buvana. 2005. Splitting the notion of 'agent': case-marking in early child Hindi. *Journal of Child Language* 32. 787-803.

Narasimhan, Buvana, Budwig, Nancy, & Murty, Lalita. 2005. Argument realization in Hindi caregiver–child discourse. *Journal of Pragmatics* 37(4). 461-495. https://doi.org/10.1016/j.pragma.2004.01.005.

Nash, John C. 2014. On Best Practice Optimization Methods in R. *Journal of Statistical Software* 60(1). 1-14. https://doi.org/10.18637/jss.v060.i02

Nash, John C., & Varadhan, Ravi. 2011. Unifying Optimization Algorithms to Aid Software System Users: Optimx for R. *Journal of Statistical Software* 43(1). 1-14. https://doi.org/10.18637/jss.v043.i09

Pienemann, Manfred. 1998. *Language processing and second language development: Processability theory* (Vol. 15). Amsterdam: Benjamins.

Ponnet, Aaricia, Kristof Baten, & Saartje Verbeke. 2016. The acquisition of differential object marking in Hindi as a foreign language. *Dutch Journal of Applied Linguistics* 5(2), 101-125. https://doi.org/10.1075/dujal.5.2.01bat.

Ponnet, Aaricia & De Cuypere, Ludovic. 2023. "Replication Data for: The acquisition of Hindi split-ergativity and Differential Object Marking by Dutch L1 speakers: systematicity and variation", https://doi.org/10.18710/3YWQ8R; DataverseNO, V1.

R Core Team. 2021. *R: The R Project for Statistical Computing*. https://www.r-project.org/.

Ranjan, Rajiv. 2016. *Acquisition of ergative case in L2 Hindi-Urdu.* The University of Iowa dissertation.

Sarkar, Deepayan. 2008. *Lattice: Multivariate Data Visualization with R*. Springer-Verlag. https://doi.org/10.1007/978-0-387-75969-2.

Spoelman, Marianne, & Verspoor, Marjolijn. 2010. Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish. *Applied Linguistics* 31(4). 532-553.

Stoll, Sabine, & Bickel, Balthazar. 2013. Capturing diversity in language acquisition research. In Balthasar Bickel, Lenore A. Grenoble, David A. Peterson, Alan Timberlake (eds.), *Language Typology and Historical Contingency: In Honor of Johanna Nichols*, 195-216. Amsterdam: Benjamins.

Tarone, Elaine. 1988. *Variation in interlanguage.* London: Edward Arnold.

Verbeke, Saartje. 2013. *Alignment and Ergativity in New Indo-Aryan Languages.* Berlin: Mouton De Gruyter.

Verbeke, Saartje & Ponnet, Aaricia. 2022. Animacy, Specificity, and Verb Semantics: What Drives Differential Object Marking in Hindi?. In Tatiana Oranskaia, Anvita Abbi (eds.), *The Heart of Change. Issues on Variation in Hindi, 133-154.* Heidelberg: HASP.

Verspoor, Marjolijn, Lowie, W., & de Bot, K. 2021. Variability as normal as apple pie. *Linguistics Vanguard*, vol. 7, no. s2, 2021, pp. 20200034. https://doi.org/10.1515/lingvan-2020-0034.

Verspoor, Marjolijn, Lowie, W., & Van Dijk, M. 2008. Variability in second language development from a dynamic systems perspective. *The Modern Language Journal* 77. 214-231.

Wickham, Hadley. 2016. *ggplot2—Elegant Graphics for Data Analysis*. Springer. https://www.springer.com/gp/book/9780387981413

Wickham, H., François, Romain, Henry, Lionel, Müller, Kirill, & Davis Vaughan. 2021. *dplyr: A Grammar of Data Manipulation (1.0.5)*. https://CRAN.R-project.org/package=dplyr.