

Linguistica e Filologia

43

Dipartimento di Lingue, Letterature e Culture Straniere
UNIVERSITÀ DEGLI STUDI DI BERGAMO 2023



BERGAMO UNIVERSITY PRESS

sestante edizioni

Direzione della rivista – Scientific Direction

Francesco Lo Monaco, Direttore responsabile/Editor in Chief Università di Bergamo

Régine Delamotte, Université de Rouen

Wolfgang Haubrichs, Universität des Saarlandes

Edgar Radtke, Universität Heidelberg

Comitato editoriale – Advisory Board

Emilia Calaresu, Università di Modena e Reggio

Luisa Chierichetti, Università di Bergamo

Silvia Dal Negro, Libera Università di Bolzano

Fulvio Ferrari, Università di Trento

Maria Pavesi, Università di Pavia

Ada Valentini, Università di Bergamo

Alessandro Zironi, Università di Bologna

Comitato Scientifico – Scientific Committee

Cecilia Andorno, Università di Torino

Alvise Andreose, Università e-Campus

Patrizia Anesa, Università di Bergamo

David Ashurst, University of Durham

Sandra Benazzo, Université de Paris VIII

Gaetano Berruto, Università di Torino

Gabriella Carobbio, Università di Bergamo

Gabriele Cocco, Università di Bergamo

Adriana Constăchescu, Universitatea din Craiova

Patrizia Giuliano, Università di Napoli ‘Federico II’

Liana Goletiani, Università di Bergamo

Roberta Grassi, Università di Bergamo

Federica Guerini, Università di Bergamo

John McKinnell, University of Durham

Giuliano Mion, Università di Cagliari

Maria Grazia Saibene, Università di Pavia

Heidi Siller-Runggaldier, Universität Innsbruck

Miriam Voghera, Università di Salerno

Marzena Wątarek, Université de Paris VIII

Maria Zaleska, Uniwersytet Warszawski

Redazione – Editorial board

Jacopo Saturno Università di Bergamo

INDICE

MARIA LAURA RESTIVO <i>I costrutti relativi con il quale e cui nel corpus UniverS-Ita</i> »	9
FEDERICA GUERINI <i>“Cose che è troppo lungo e pericoloso scrivere”.</i> <i>Note sull’uso del bergamasco con funzione criptica</i> <i>nelle lettere di Papa Giovanni XXIII</i> »	43
TERESA CARBUTTI <i>Il Raddoppiamento Fonosintattico come marca di genere</i> <i>nei dialetti lucani</i> »	75
ORESTE FLOQUET <i>Apport d’un sondage d’opinion à l’étude du gérondif négatif ..</i> »	101
LUISA CHIERICHETTI <i>Reescrituras telecinemáticas:</i> <i>transformaciones textuales en la serie Patria</i> »	123
ANGELA ANDREANI, DANIEL RUSSO <i>Building a Corpus of the Metalanguage</i> <i>of English Linguistics 1500-1700: Methodological Issues</i> »	151
MATTEO DE FRANCO <i>Studi sul lessico dello Statuto della gilda</i> <i>di Santa Caterina in gultnico antico.</i> <i>Una nuova interpretazione del lessema abita</i> »	175
ANNA BONOLA, VALENTINA NOSEDA <i>La violazione delle presupposizioni</i> <i>nei discorsi sul Russkij mir</i> »	195

Indice

AARICIA PONNET, LUDOVIC DE CUYPERE
*The acquisition of Differential Object Marking:
a longitudinal study on L1 Dutch learners
of Hindi as a foreign language* » 217

Linguistica e Filologia

43

Linguistica e Filologia è inclusa in ERIH PLUS
(European Reference Index for the Humanities and Social Sciences)

Internet: <http://aisberg.unibg.it/handle/10446/6133>

I contributi contenuti nella rivista sono indicizzati nelle banche dati
Modern Language Association (MLA) International Bibliography
e Linguistics and Language Behaviour Abstracts (LLBA),
Directory of Open Access Journals (DOAJ) e Web of Science

Licenza Creative Commons:

This journal is published in Open Access under a Creative Commons License
Attribution-Noncommercial-No Derivative Works (CC BY-NC-ND 3.0).

You are free to share – copy, distribute and transmit –
the work under the following conditions:

You must attribute the work in the manner specified by the author or licensor
(but not in any way that suggests that they endorse you or your use of the work).

You may not use this work for commercial purposes.

You may not alter, transform, or build upon this work.



Volume pubblicato dal Dipartimento di Lingue, Letterature e
Culture Straniere e finanziato con fondi di Ateneo di ricerca.

ISSN: 1594-6517

AARICIA PONNET
(Ghent University)

LUDOVIC DE CUYPERE
(Ghent University/Vrije Universiteit Brussel)

The acquisition of Differential Object Marking: a longitudinal study on L1 Dutch learners of Hindi as a foreign language

Abstract

This article investigates the acquisition of Differential Object Marking (DOM) in Hindi as a foreign language. The study aims to verify earlier cross-sectional studies on Hindi DOM, and their earlier explanations related to L1 influence, item-based learning and semantic mapping. Via multiple oral interviews with five L1 Dutch speakers we compiled a longitudinal learner corpus of $n = 3684$ utterances. Our findings confirm high omission rates, suggesting initial expansion of the optionality of Hindi DOM (possibly enhanced by L1 transfer) and late acquisition of functional DOM use (possibly via item-based learning).

Keywords: Differential Object Marking, Hindi, Syntax-semantics interface, Foreign Language Acquisition, Learner corpus data

1. Introduction

This article investigates the development of Differential Object Marking (DOM) with Dutch learners of Hindi. The acquisition of DOM has gained interest among researchers studying different languages such as Spanish, Romanian, Basque and Hindi (Mardale & Montrul 2020). DOM is a common phenomenon whereby the Direct Object (DO) is morphologically marked depending on features such as animacy, specificity, definiteness, and/or referentiality (Malchukov 2008). DOM manifests itself in multiple ways cross-linguistically. In Hindi, DOs receive the DOM marker *-ko* depending on animacy and specificity. Research on Hindi DOM has focused on its use and acquisition by heritage speakers in the USA (Montrul et al. 2012; 2015; 2019), English-speaking foreign language learners (Montrul et al. 2019) and Dutch-speaking foreign language learners (Ponnet et al. 2016; Ponnet & De Cuypere 2023a). A recurrent finding is that learners have difficulties using

DOM and omit the *-ko* marker. Moreover, a longitudinal study with child L1 Hindi learners reports on similar findings (Narasimhan 2005; 2020).

The present study investigates the development of DOM with Dutch-speaking learners of Hindi as a Foreign Language (HFL). Dutch does not overtly mark the DO and (although it generally is analysed as an SOV language (Vandeweghe 2007: 234)) has SVO word order in the main clause (Koster 1975; Zwart 1997).

Taking a Multiple Case Study approach, we compiled a learner corpus of $n = 3684$ utterances with learners who were enrolled in the Hindi language program at a higher education institute in Belgium. We analysed the learner language on the basis of four semi-spontaneous interviews eliciting DOM, carried out in the course of the participants' first and second year of study.

The first aim of the study is to examine how DOM develops longitudinally in the interlanguage of L1 Dutch-speaking HFL learners. So far, the acquisition of HFL has been examined cross-sectionally, which is a methodologically suboptimal approach. By means of our longitudinal investigation we wish to verify earlier findings in cross-sectional studies on HFL.

The second aim is to examine three existing explanations for the acquisition of Hindi DOM, including: (i) L1 influence (Montrul et al. 2012; 2015; 2019), (ii) item-based learning (Narasimhan 2005; 2020) or (iii) semantic mapping (Baten & Ponnet 2023).

The article is organized as follows. Section 2 provides an overview of Hindi DOM and addresses the differences in the way the acquisition of DOM in Hindi has been researched and theoretically motivated. Section 3 provides details on the background of the learners, outlines the collected corpus data and describes the methodology. Section 4 presents the results of the quantitative analysis. Section 5 discusses the results and adds some qualitative remarks to the analysis. Section 6 rounds off the article and offers some suggestions for future research.

2. The acquisition of Hindi DOM

2.1 Overview Hindi DOM

Hindi DOM is governed by the animacy and specificity of the Direct Object (DO) (Mohanani, 1994). As summarized in Table 1, animate DOs are marked with *-ko*, whereas inanimate objects are only marked with

-ko when the DO is specific. Additionally, *-ko* marks other semantic roles such as the Recipient, the Experiencer, whereby *-ko* marking is always obligatory, and Goals, where marking is also optional (Table 1). In this article, we will refer to the marking of the DO, the Recipient, the Experiencer and the Goal as *-ko* marking:

Table 1: *-ko* marking in Hindi¹

-<i>Ko</i> marking of DOM	
Animate	
• Specific	<i>-ko</i>
• Non-specific	<i>-ko</i> [Ø] ²
Inanimate	
• Specific	Ø/ <i>-ko</i>
• Non-specific	Ø
-<i>Ko</i> marking of other roles	
Recipient	<i>-ko</i>
Experiencer	<i>-ko</i>
Goal	Ø / <i>-ko</i>

2.2 *Transfer; item-based learning and semantic mapping*

A handful of studies have been written on the acquisition of case marking in Hindi. Interestingly, all of the studies report high omission rates for Hindi DOM, even though they investigate different speaker populations and give different theoretical explanations for these omission rates. All of the studies on heritage speakers (Montrul et al. 2012; 2015; 2019) and FL acquisition (Baten & Verbeke 2015; Montrul et al. 2019; Ponnet et al. 2016; Ponnet & De Cuypere 2023a) are cross-sectional studies. Since these studies report on similar findings, perhaps some general linguistic dynamics related to the nature of DOM are at play. We therefore aim to test different explanations provided in the literature for the acquisition of DOM

¹ Please see appendix for a more in-depth overview of *-ko* marking in Hindi.

² Exceptionally, in a few instances with animate, non-specific DOs that render a more generic meaning, the zero-morph may be found rather than *-ko*. See appendix.

in Hindi, and also include in our overview a longitudinal study with child learners of Hindi (Narasimhan 2005; 2020) and a theoretical contribution on the acquisition of DOM (Baten & Ponnet 2023).

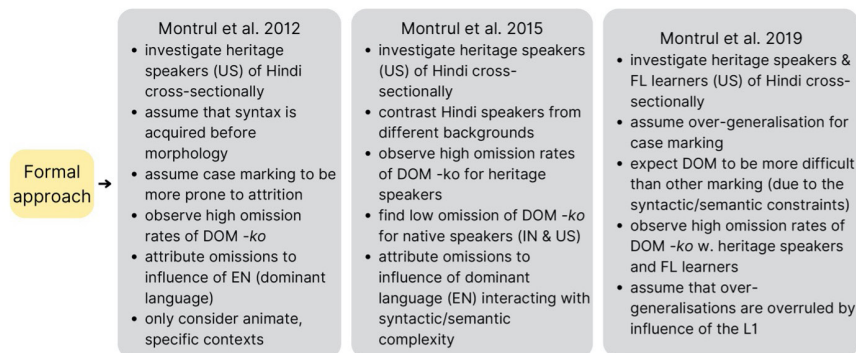
The studies on heritage speakers in the US describe significantly high rates of omission in comparison with native speakers in India. The three studies by Montrul et al. (2012; 2015 and 2019) build on findings in studies on other DOM languages such as Spanish and Romanian where it has been shown that DOM is particularly prone to attrition. Montrul et al. (2019) additionally investigates a group of intermediate and advanced FL learners of Hindi. The three studies take a formal approach, and especially Montrul et al. (2012) and Montrul et al. (2019) hypothesize that there is a set of rules and principles that allows them to predict the use of *-ko* marking. All three studies find high omission rates of *-ko* with second generation Hindi speakers and FL learners in the US, and attribute the omission of *-ko* to: 1) the complex form-meaning mappings that make DOM *-ko* more difficult than other functions of *-ko*; 2) influence of the dominant language, English, that does not have overt case marking. Important to note is that Montrul et al. (2012) and Montrul et al. (2019) only consider DOs that are both animate and specific in their analysis of the oral production data. Interestingly, Montrul et al. (2015) do not find high omission rates with native speakers in India or the US, neither do they find overgeneralisations, suggesting that DOM is not grammaticalizing like in Spanish where speakers of certain varieties either expand the use of DOM, or share the preference for discarding the use of DOM (García 2018; Bautista-Maldonado & Montrul 2019).

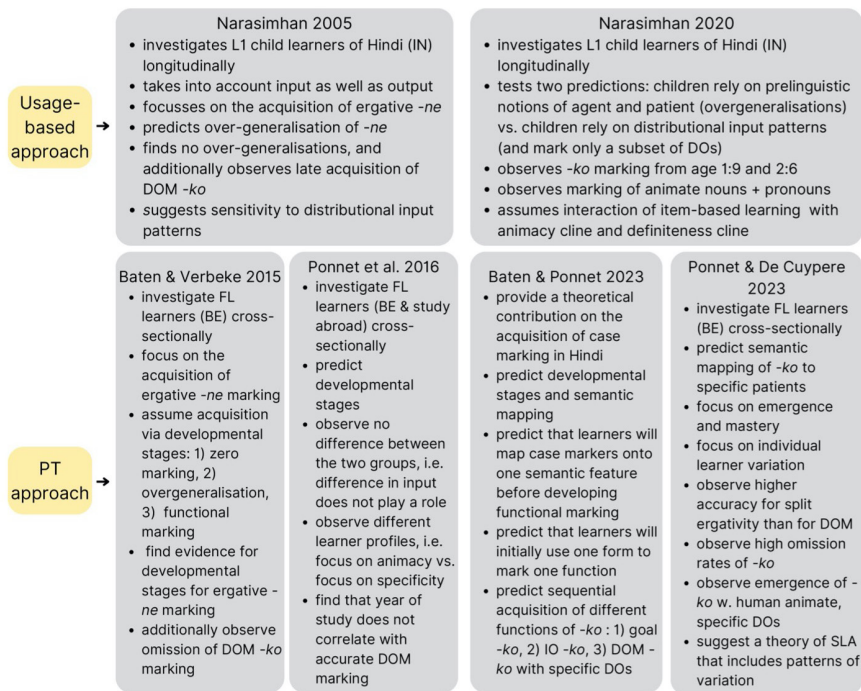
Narasimhan (2005; 2020) performed a longitudinal study with child learners of Hindi in India. In the first study, Narasimhan (2005) only provides preliminary results on the acquisition of DOM. The three child learners already use the ergative marker *-ne* but barely use *-ko* as a DO marker at all (and only in sentences where *-ne* is also used). Narasimhan (2020) finds late emergence (after age 1:9 and 2:6) and omission of *-ko* with two child learners of Hindi as well, but notices a pattern in their marking behaviour: the children only mark animate DOs and pronouns. Narasimhan focuses on both the input and output of the child learners, and advances a usage-based approach in order to explain the data: she proposes acquisition of case forms through frequency in the input, as a kind of item-based learning where the use of case markers is linked to certain learned verbs and contexts.

Three empirical studies and one theoretical contribution have been published on the acquisition of DOM in Hindi from a Processability Theory (PT) approach (Pienemann 1998). These studies aim to predict and explain the acquisition of DOM through emergence and developmental stages. One study focusses on the acquisition of ergative *-ne* (Baten & Verbeke 2015) with Dutch HFL learners and finds omission of DOM *-ko* in their data. A study with learners of Hindi in Belgium and India (study abroad) from different language backgrounds found high omission rates of *-ko* as well. The authors did not find solid evidence for developmental stages but instead discerned different learner profiles based on the semantic-syntactic constraints the learners paid attention to. Baten & Ponnet (2023) suggest that Hindi DOM develops through semantic mapping, whereby HFL learners focus on the feature specificity before they acquire functional DOM marking and apply the case markers according to both the features animacy and specificity. This has been tested in a cross-sectional study with Dutch learners by Ponnet & De Cuypere (2023a), who falsified this assumption and found that DOM *-ko* emerges with human animate, specific DOs, but the authors again found high rates of omission with beginning and advanced learners of Hindi. Additionally, they observe high accuracy rates for other case markers with advanced learners and high interlearner variation.

Table 2 provides an overview of previous scholarship on the acquisition of Hindi DOM in relation to their theoretical framework:

Table 2: overview theoretical frameworks DOM acquisition





Based on the overview of the different explanations for the acquisition of DOM in Hindi in Table 2, we aim to test the following expectations:

- If, like Montrul et al. (2019) suggest in their study on balanced and unbalanced bilinguals (including intermediate and advanced foreign language learners), the presence of the features that require *-ko* marking is overruled by the dominant language (in this case, Dutch), then we should find omission rates of *-ko* to a high extent depending on the learners’ proficiency, and thus possibly also towards the end of our longitudinal study.
- If the learners show sensitivity to the syntactic-semantic constraints that guide the use of *-ko* marking, the question arises as to how this develops into functional case marking, i.e. accurate *-ko* marking according to the conditions animacy and specificity. Narasimhan (2005; 2020) suggests that learners might expand their use of case marking by means of item-based learning. If this is the case, then

- we should find the use of *-ko* only with specific verbs and/or nouns, and then we should find *-ko* marking with different types of DOs regardless of the features animacy or specificity.
- If learners use a strategy called semantic mapping (i.e. they map the use of *-ko* onto one semantic feature regardless of the functional restrictions of DOM) and focus on either one of the semantic-syntactic constraints (cf. Baten & Ponnet 2023; Ponnet & De Cuypere 2023a), we should find *-ko* to emerge with animate specific DOs during the first observations, only to find *-ko* marked inanimate specific DOs towards the last observations of the longitudinal study.

3. Methodology

We used a picture description task to elicit semi-spontaneous oral production data from $N = 5$ learners of Hindi in Belgium. Following a longitudinal design, we gathered data during four observations spread over three semesters. These took place in the second (Time 1), third (Time 2 and 3) and fourth semester (Time 4) of the learners' Hindi course trajectory. The reason why we focus on spoken language production is that we wish to investigate the HFL learners' active use of *-ko* marking.

3.1. Participants

We recruited five Dutch-speaking students of Hindi, all enrolled in a full-time course on Indian languages and cultures at a Belgian university. The Hindi language course was their major subject and their second major was a Sanskrit language course. All participants were female, Flemish, white, and high school educated. We have no information about their socio-economic background but asked them to fill in a questionnaire with some questions regarding out of class activities and other known languages³. At the start of data elicitation the students were 18-19 years old, at the end of data elicitation 19-20. The students were recruited on a voluntary basis, signed an informed consent form and received neither study credits nor money for their participation in the study. The students were motivated to participate in the study and stated that they appreciated the opportunity to practice their oral language skills. The socio-linguistic

3 This information is accessible via Ponnet & De Cuypere (2023b).

background questionnaire also revealed learner motivation (mostly an interest in India, its culture, and the Hindi language itself) and out-of-class activities such as watching Hindi movies and chatting with friends.

3.2. *Materials and procedure*

We designed a picture description task with illustrated pictures adapted from a comic book (Dulieu, 1972; 1974) and self-created pictures. Two different sets of pictures with similar actions were created, which we refer to as task 1 and task 2 in table 3.

Table 3: oral production task

Task 1	Narrative with comic book pictures: 17 slides with 30 pictures. Narrative with self-created pictures: 25 slides with 51 pictures.
Task 2	Narrative with comic book pictures: 19 slides, 37 pictures. Narrative with self-created pictures: 23 slides, 41 pictures.

We alternated between task 1 and task 2 for the four different data collection points (cf. variable Time, further outlined below). As this research is aimed at eliciting sentences with different types of DOs, pictures were ordered in a narrative structure, similar to a story-telling task. In this way, we aimed at providing sufficient contexts where a DO should be produced. By introducing new characters into the storyline, we could alternate between specific and non-specific referents. To validate our elicitation task, i.e., to ensure that our task elicited DOs, we piloted the task with five native speakers. This confirmed that the task was fit to elicit data for DOM and other functions of *-ko*.

The participants were told they were partaking in a study that investigated the development of Hindi, without the specific focus of this study, i.e. the acquisition of DOM, being disclosed. A debrief at the end of the study confirmed that the students were not aware of the specific focus of the study, most stated that they thought the research was about sentence production. During the interview, the students were provided with a vocabulary list (see appendix) that they could freely consult during the task. The participants were allowed to speak in a self-paced way, they were only encouraged to continue after longer silences by cues such as

aur ‘and’ and *phir* ‘and then’. A specific set of questions was designed to make sure the students produced enough sentences with a DO. The interviewer would conduct the interview in Hindi, only switching to Dutch if the student did not understand.

3.3. *Timing interviews and instruction/input*

The students were interviewed during four data collection points, one at the end of their first year of study, and three during their second year of study. In total the participants received 24 weeks of instruction per year, spread over two semesters (September-December and February-May).

Table 4: longitudinal study points in time vs. amount of instruction

	Longitudinal study points in time	Weeks of instruction	Total hours of instruction (Hours of instruction/ week)
Time 1	Halfway 2 nd semester Y1	ca. 20 weeks	90 – 120h (4,5 – 6h Hindi/week)
Time 2	End 1 st semester Y2	ca. 32-34 weeks	144 – 153h (4,5h Hindi/week)
Time 3	Beginning 2 nd semester Y2	ca. 40 weeks	180h (4,5h Hindi/week)
Time 4	End 2 nd semester Y2	ca. 48 weeks	216h (4,5h Hindi/week)

Table 4 connects the data collection points with the amount of instruction the participants received. The timing of the data collection points is motivated as follows: the students were introduced to the DO marker *-ko* only in their 2nd semester of their Hindi study. The 2nd year of the Hindi course is an intense trajectory that prepares the students for their study abroad in year 3, this is why we decided to interview the students three times during year 2, as we expected more acquisition dynamics within the interlanguage of the learners as they had already had a full year of input and would receive more instructed knowledge in the second year.

Regarding input, four out of the five learners (Learner a, b, d and e) spent a month in India between Time 1 and Time 2, stating that they interacted with native speakers. We checked the learner materials that the learners had access to and it is worth mentioning that Hindi teaching materials currently discuss DOM only in a brief manner. We noticed that the instruction on DOM *-ko* marking in English teaching materials on Hindi as well as in the teaching in class is rather limited (especially when compared to the explicit instruction on ergative *-ne* marking and Recipient/Experiencer *-ko* marking). Moreover, the multiple functions of *-ko* are all introduced in the first year. DOM is included in the learner materials implicitly from the beginning of the second semester and is the final function of *-ko* to be introduced explicitly, halfway the second semester (correlating with the first observation point, Time 1).

3.4. *Statistical data analysis*

The interviews were transcribed by the first author of this article and by an MA student who had finished their Hindi course (not one of the students participating in the study). The transcriptions were exported as a .txt file and the utterances were annotated in MS Excel.⁴ For both DOM *-ko* and the other uses of *-ko*, we first performed a bivariate analysis looking at observed frequencies and proportions of each predictor variable in relation to the use of *-ko*. Then we fitted a mixed-effects logistic regression model, starting with a full model with random intercepts and random slopes and interactions with Time. The random slope was dropped in both models (DOM *-ko* and other *-ko*) due to data sparsity. Further modelling details are provided in the results section.

To evaluate the Item-based learning hypothesis, we created bivariate tables in which the most frequent Items, i.e., the head Noun of the DO Noun Phrase (DO_Item), the sentence verb (V_Item), or the head noun of the Noun Phrase that can take non-DOM *-ko* marking (I_Item), were tabulated against the use of *-ko*.

All statistical analyses were performed with R (R Core Team 2021) in RStudio (RStudio Team 2022). We used the following packages (in alphabetical order): dplyr (Wickham et al. 2022), ggplot2 (Wickham

4 For information about the annotated variables, see Appendix.

2016), ggeffects (Lüdtke 2018), Hmisc (Harrell 2022), lme4 (Bates et al. 2015), sjPlot (Lüdtke 2022). A notebook with the full data analysis and output is shared in Ponnet & De Cuypere (2023b).

4. Results DOM

In total, we elicited more than 13 hours of semi-spontaneous oral production data on the basis of which we compiled a dataset of $n = 3684$ utterances. There were 1522 (41%) clauses with a Direct Object (DO), that could potentially take a *-ko* marker. Of these 1522 DOs, 120 (7%) were marked with *-ko*.

Table 5 presents the bivariate summary statistics of the use of *-ko* in relation to Time, DO specificity and Animacy, and Learner. The descriptive statistics show that the use of *-ko* increases with time and that *-ko* is only once used with a non-specific DO. The DOM marker appears to be used with animate DOs. Learners a and b seemed to have used *-ko* more often than the other learners.

Table 5: Descriptive statistics for DOM in relation to Time, DO specificity, DO Animacy, and Learner.

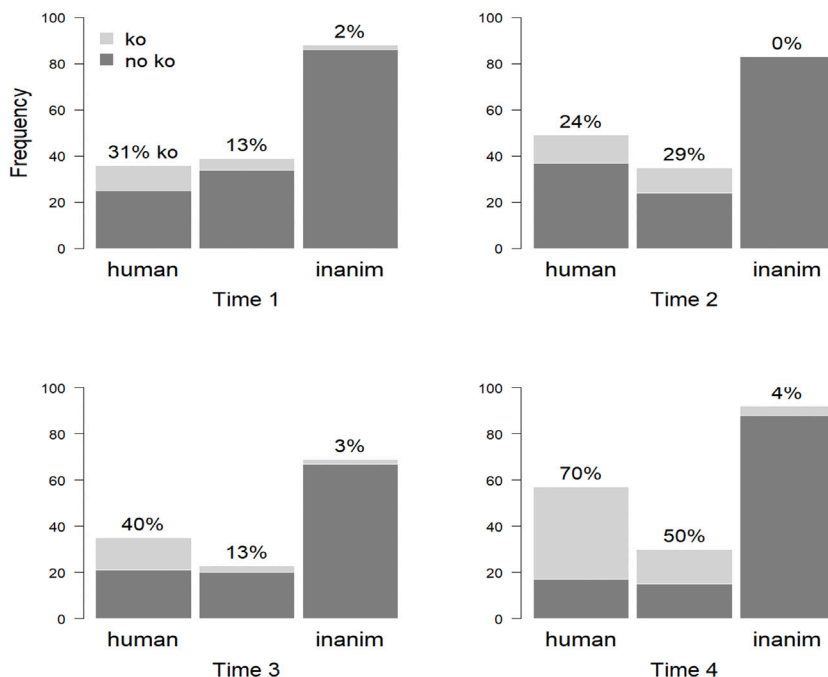
	DOM	
	ko	no ko
Time 1	18 (5%)	376 (95%)
2	23 (6%)	364 (94%)
3	20 (6%)	342 (94%)
4	59 (16%)	321 (84%)
DO Specificity		
specific	119 (19%)	518 (81%)
non-specific	1 (0%)	885 (100%)
DO Animacy		
human animate	77 (27%)	204 (73%)
non-human animate	35 (10%)	333 (90%)
inanimate	8 (1%)	866 (99%)

Learner

a	32 (10%)	277 (90%)
b	44 (13%)	282 (87%)
c	22 (6%)	329 (94%)
d	4 (2%)	261 (98%)
e	18 (7%)	254 (93%)

Figure 1 shows the use of *-ko* in relation to DO Animacy in the course of the 12 months of observations, restricting the data to specific DOs only (recall that only one non-specific DO was used with *-ko*). The overall majority of elicited DOs is inanimate, which seldom receive *-ko* marking. We also observe an increase in the use of *-ko* for both human and non-human animate DOs.

Figure 1: The use of *-ko* in relation to DO Animacy by Time.



We evaluated the multivariate effect of DO Animacy, Time and Learner by means of a mixed-effects logistic regression model. The data were again restricted to specific uses only, given that no *-ko* marking is used with non-specific DOs. The random slope term for Time by Learner was omitted because of lack of data (we received a singularity warning, which suggests overfitting). An interaction effect between DO Animacy and Time was not significant based on a Likelihood Ratio Test of nested models ($X^2 = 9.9$, $df = 6$, $P\text{-value} = 0.13$) and was accordingly dropped. The models estimates of our final model are given in Table 6:

Table 6: Estimates for the mixed-effects logistic regression model. The use of -ko is treated as the success. DO animacy is dummy coded, with DO animate as the reference level. For Time, we use forward difference coding, so every level is compared to the next one.

Fixed effects	Estimate (SE)	Z	P-value
Intercept	-1.55 (0.57)	-1.0	0.33
Time 1-2	-0.37 (0.39)	-0.9	0.35
Time 2-3	-0.53 (0.40)	-1.3	0.18
Time 3-4	-1.41 (0.38)	-3.7	0.002
DO inanim	-4.12 (0.44)	-9.4	< 0.001
DO non-human	-0.86 (0.29)	-2.9	0.003
Random Intercept	Var	SD	
Learner	1.43	1.20	

The two significant main effects for Time and DO Animacy are visualised by means of their predicted probabilities in Figure 2 and can be interpreted as follows. There is an increase of the use of *-ko* over Time and *-ko* is significantly more often used with human and non-human animate DOs than with inanimate DOs. The increase from Time 3 to Time 4 is statistically significant ($P\text{-value} = 0.002$). Although the interaction effect between Time and DO Animacy was not significant, there is a notable change in the effect of DO Animacy over Time, in that the effect for Animacy becomes more outspoken. More specifically, at Time 4, the predicted probabilities for *-ko* are clearly higher for human and non-human DOs than for inanimate DOs than at Time 1.

Figure 2: Predicted probabilities for *-ko* conditional on Time and DO Animacy (two combined main effects).

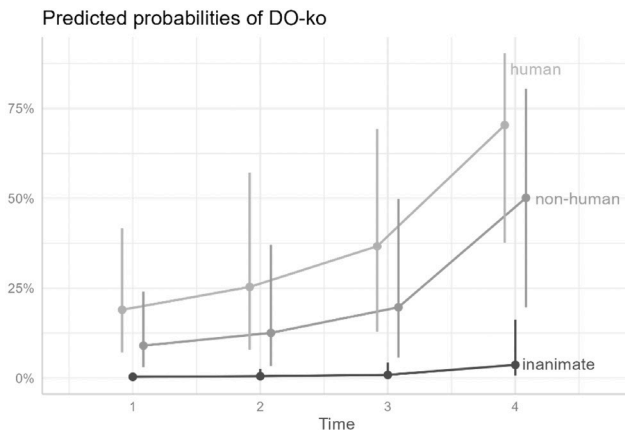
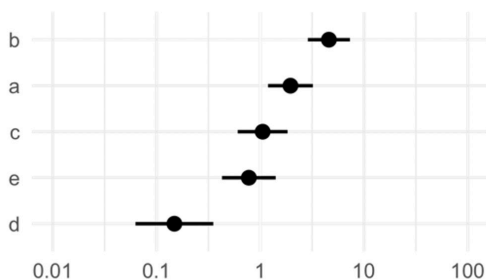


Figure 3 shows further that Learner b tends to use *-ko* more often than average, while the opposite is true for Learner d.

Figure 3: Conditional modes for Learner (on the odds scale).



Looking at the use of *-ko* by the Learners over Time in Table 7, we can see that there is a noticeable increase from Time 3 to 4 for Learners a and e, whereas Learner b uses *-ko* more often than the other Learners already in Time 1 and 2.

*Table 7: Use of -ko by Learner over Time. *DO -ko– DO Ø.*

Learner	Time			
	1	2	3	4
a	3 – 94*	6 – 63	7 – 60	16 – 60
b	11 – 84	12 – 84	5 – 60	16 – 54
c	1 – 87	3 – 67	6 – 90	12 – 84
d	3 – 50	0 – 74	1 – 67	0 – 70
e	0 – 61	2 – 76	1 – 64	15 – 53

Additionally, we also performed an analysis to investigate whether we can observe item-based DOM *-ko* marking in our data (see Appendix for the results and tables).

5. *-Ko marking of the Recipient, Experiencer and Goal*

We elicited N = 394 utterances which could also be marked with *-ko*, namely Recipients, Experiencer subjects and Goals.

Table 8 shows an increase in the use of *-ko* from in Times 3 and a particular preference for *-ko* marking with the Semantic Roles of Experiencer and Recipient.

Table 8: Descriptive statistics for the other uses of -ko in relation to Time, Semantic Role and Learner.

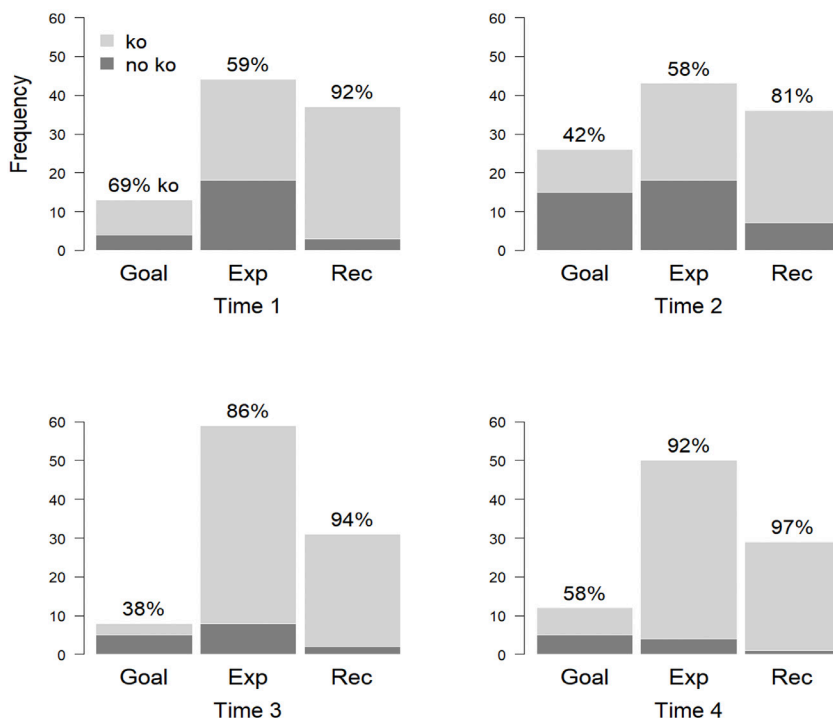
	Ko	
	Yes	No
Time 1	69 (73%)	25 (27%)
2	65 (62%)	40 (38%)
3	83 (85%)	15 (15%)
4	81 (89%)	10 (11%)
Semantic Role		
Goal	30 (51%)	29 (49%)
Experiencer	148 (76%)	48 (24%)
Recipient	120 (90%)	13 (10%)

Participant

a	68 (92%)	6 (8%)
b	86 (77%)	25 (23%)
c	36 (63%)	21 (37%)
d	63 (72%)	25 (28%)
e	45 (78%)	13 (22%)

Figure 4 shows an increase in the use of *-ko* as a marker of Experiencer and Recipient over Time. The use of *-ko* marking for the latter roles is particularly high in Time 4, both Experiencers and Recipients receive *-ko* marking in over 90% of all cases. The longitudinal trend of *-ko* as a Goal is more variable. There is actually a decrease from 69% in Time 1 to 58% in Time 4.

Figure 4: the use of non-DOM-*ko* in relation to Semantic Role over Time.



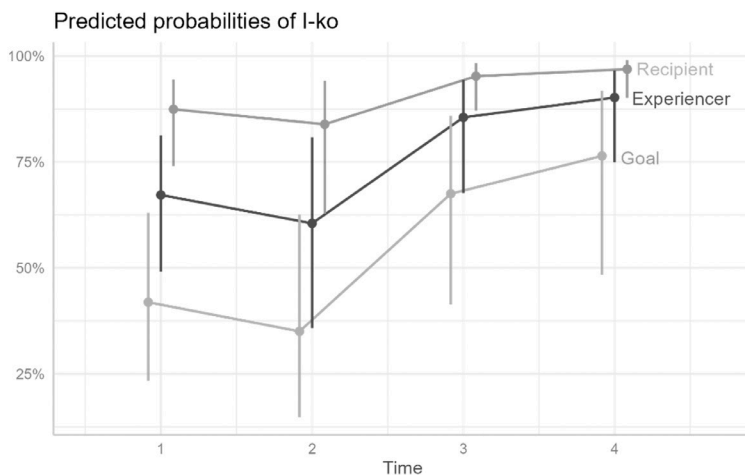
Based on a mixed-effects logistic regression analysis, we found evidence for the main effects of Time and Semantic role. The random slope for Time by Learner was again not retained and no evidence for an interaction effect between Time and Semantic role was found based on a Likelihood Ratio Test ($X^2 = 9.7$, $df = 6$, $P\text{-value} = 0.13$). We included Learner as a random intercept to account for clustering. The model estimates are given in Table 9:

Table 9: Estimates for the mixed-effects logistic regression model. The use of -ko is treated as the success. Semantic Role is dummy coded with Goal as the reference level. For Time, we use forward difference coding, which means that every level is compared to the next one.

Fixed effects	Estimate (SE)	Z	P-value
Intercept	0.25 (0.52)	-0.6	0.56
Time			
1-2	0.28 (0.34)	0.8	0.404
2-3	-1.43 (0.39)	-3.7	< 0.001
3-4	-0.37 (0.46)	-0.8	0.418
Semantic Role			
Experiencer	1.04 (0.35)	3.0	0.002
Recipient	2.34 (0.42)	5.5	< 0.001
Random Intercept	Var	SD	
Learner	0.52	0.72	

The main effects for Time and Semantic Role can be interpreted by means of the effect plot in Figure 5. The use of *-ko* significantly increases from Time 2 to 3, after which it seems to stabilize towards Time 4. There is also a small and non-significant drop from Time 1 to Time 2. The use of *-ko* with a Recipient and Experiencer is significantly higher than with a Goal irrespective of Time. Although no evidence was found of an interaction between Time and Semantic Role, there is a notable increase in the use of *-ko* for Goal and Experiencer from Time 1 to Time 4.

Figure 5: Predicted probabilities for non-DOM *-ko* conditional on Time and DO Animacy (two combined main effects).



The random intercept predictions further show that Learner a uses non-DOM *-ko* more often than the average Learner. The opposite is found for Learners d and c (Figure 6).

Figure 6: Conditional modes for the random intercept of Learner.

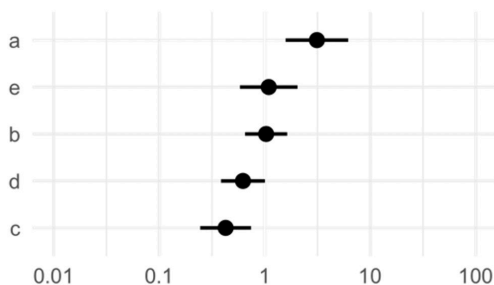


Table 10 illustrates that Learners a, b and d are the most frequent users of non-DOM *-ko* marking.

Table 10: Use of *-ko* for Recipient, Experiencer and Goal by Learner over Time. **-ko*– \emptyset .

	Learner	Time			
		1	2	3	
a	Recipient	9 - 0	8 - 0	3 - 1	5 - 0
	Experiencer	11 - 0	7 - 0	8 - 0	11 - 0
	Goal	5 - 2	1 - 3	0 - 0	0 - 0
b	Recipient	13 - 1	9 - 0	12 - 0	6 - 0
	Experiencer	5 - 11	6 - 8	13 - 0	8 - 2
	Goal	4 - 0	4 - 2	1 - 0	5 - 1
c	Recipient	3 - 1	2 - 4	5 - 1	4 - 0
	Experiencer	6 - 2	1 - 2	6 - 3	9 - 0
	Goal	0 - 0	0 - 5	0 - 4	0 - 0
d	Recipient	6 - 0	3 - 3	6 - 1	5 - 1
	Experiencer	2 - 5	8 - 4	21 - 0	11 - 2
	Goal	0 - 1	0 - 3	1 - 1	0 - 4
e	Recipient	3 - 1	7 - 0	3 - 0	8 - 0
	Experiencer	2 - 0	3 - 4	3 - 5	7 - 0
	Goal	0 - 1	6 - 2	1 - 0	2 - 0

6. Discussion

From the results section above, we can remember the following important points. Firstly, *-ko* emerges relatively early as a DO marker in the interlanguage of the learners in our study. Four out of five learners already use *-ko* as a DO marker at Time 1. Secondly, *-ko* is only used with specific (identifiable) DOs in our dataset. Already at Time 1, we find *-ko* marking of the following DOs: [+human animate; specific], [non-human animate; specific]; [inanimate; specific]. Thirdly, there is a clear association of *-ko* with specific, human animate DOs from Time 1 onwards, which significantly increases towards Time 4. The data also show that *-ko* marking gradually expands throughout the different data collection points to specific non-human animate DOs and specific

inanimate DOs, and shows a significant increase for these types of DO between Time 3 and Time 4 as well. However, marking of specific inanimate DOs remains relatively low in comparison to marking of both human animates and non-human animates. This suggests that *-ko* marking emerges when both the factors specificity and animacy coincide.

To control for the multifunctionality of the Hindi case marker *-ko*, we also included the following functions of *-ko* in our analysis: Recipient, Experiencer and Goal. Recall that for Recipients and Experiencers the use of *-ko* is mandatory, for Goal the use of *-ko* is optional. The data of the use of *-ko* with these functions shows a high frequency of use already at Time 1 in comparison with the data for DOM *-ko* marking. For all three functions, we attested a significant increase between Time 2 and Time 3. We find the highest rates of accuracy for the Recipient, for which we observe an accuracy rate of more than 90% at Time 1 already, a trend which is maintained at Time 3 and Time 4 (with a small but insignificant drop in Time 2). For the Experiencer, we observe a different pattern. At Time 1, learners still tend to make quite some errors of omission with an accuracy rate of 59%. The increase between Time 2 and Time 3 is the most prominent for the Experiencer role, increasing from 58% to 86%. The accuracy rates for the Goal are more variable, which resembles the variability we can see for DO *-ko* marking, where marking of certain DOs is also optional.

Finally, in the conditional modes we observe great interlearner variability, with Learners a and b frequently using the *-ko* marker for both DOM and other functions of *-ko* and particularly Learner d having very low accuracy rates for case marking. For Learner d, who only sporadically uses *-ko* as a DOM marker and eventually does not develop functional marking, not even for human animate specific DOs, we assume a state of fossilization of DOM. Learner e, on the other hand, shows a similar marking pattern to Learner d from Time 1 to Time 3, but at Time 4 seems to have identified the context for *-ko* marking and suddenly shows an overall increase of *-ko* marking, also for the other roles. This aligns with the results for Learner a, b and c. Learner b regularly uses *-ko* for the different roles from Time 1 onwards, and continues to do so, with some range of variability, at Time 2 to Time 4. Learner a and c experience more competition between *-ko* marking of the experiencer roles and other functions, but at Time 4 they use the *-ko*

marker regularly for all of the functions.⁵ Based on our data, we thus assume that learners who performed well for case marking tended to continue performing well, whereas learners who started off with low accuracy rates for *-ko* marking also showed low *-ko* marking at the end of the study. Time spent in India between Time 1 and Time 2 did not affect DOM acquisition (this correlates with findings by Ponnet et al. (2016) who found no difference in DOM proficiency between a group studying in Belgium and a study abroad group), nor did self-reported amount of out-of-class activities (learner d was the student who reported the highest rate of out-of-class activities). Learners b and c previously learnt German, a case language, and Learner d and e learnt Spanish, a DOM language, which given the high rates of omission does not necessarily result in an advantage – but more transfer-focused research is needed to make solid claims about this.

This brings us to the initial aims of our study. Our first aim was to examine how DOM develops longitudinally in the interlanguage of L1 Dutch-speaking HFL learners, and, in particular, to verify earlier findings in cross-sectional studies on L2 Hindi and Hindi as a heritage language. Similar to the other studies on the acquisition of DOM in Hindi, we observe high omission rates for DOM *-ko* – and this throughout the entire length of the study. That we find high omission rates at Time 1 is not so surprising, at this point learners have had only one year of Hindi instruction and their interlanguage is still limited – the low accuracy rates for *-ko* marking of the Experiencer confirm this. What is noteworthy is that the omission rates for DOM remain high, even though the interlanguage of the learners has developed further. Consider the following examples (1) and (2) by Learner d (the learner who realised only four *-ko* marked DOs):

5 Except for the Goal, which seems to have a different route of development altogether. This function of *-ko* also allows for dropping of the case marker, and in a lot of instances it could even be regarded more native-like to not use the *-ko* marker as a marker of direction or time, although it is usually one of the first functions with which the use of *-ko* is introduced in learner materials. We may assume a more regular use of the case marking postposition *-ko* as a marker of direction or time in initial stages of the learner language, which eventually decreases once the interlanguage further develops. However, more data on this function of *-ko* is needed (these qualitative assumptions are based on a rather low number of utterances for this specific function).

- (1) *baunā.ø kuch.ø dekh rahā thā* [d 1458]
 Gnome.NOM something.NOM to see.PROG.PST
 ‘The gnome was seeing something.’
- (2) *to us=ne āvāz.ø sunkar* [d 5226]
 So he=ERG sound.NOM to hear.ABS
pīche dekhā hai
 behind to look.PERF.PRS
 ‘So after hearing a sound he has looked behind [him].’

Both sentences describe a similar action of the oral production task, however the first example was elicited at Time 1 and the second example at Time 4. Even though the interlanguage of Learner d at Time 1 has already reached a certain level of proficiency, at Time 4 the learner clearly has more tools to express what she is seeing: notice the use of the ergative case marker, the perfective verb form and the absolutive in the second sentence. The performance of the learners for DOM is interesting exactly because they perform quite well on the Hindi elicitation task in general. Consequently, a great challenge of our study is that the number of attestations of *-ko* marked DOs is very low, even though in total we elicited a large number of different utterances with different DO types – and our task provided multiple contexts where the use of *-ko* was required. The question then remains: how can we account for this tendency of case marking, and especially, how can we account for the high rate of omissions? This brings us to the second aim of our study, i.e. finding evidence towards one of the theoretical assumptions that we have described above: L1 influence, item-based learning or semantic mapping.

We provided an overview of the studies that have hitherto discussed and explained the acquisition of DOM in Hindi. Montrul et al. (2012; 2015; 2019) explain the high omission rates in Hindi DOM by focusing on the fact that English does not have overt case marking. Regarding heritage learners, this means English as the dominant language with unbalanced bilinguals, and English as the L1 for FL learners. Other research on DOM with learners of Turkish and Romanian (Montrul & Gürel 2015; Montrul 2019) suggests that positive transfer between DOM languages occurs. Our data indeed confirm high omission rates which suggest a role for transfer, however, L1 influence or transfer remains

a difficult topic in SLA research. L1 influence may partly account for the high omission rates of DO *-ko* in our data, but we cannot assume this without certain reservations. The learners do show *-ko* marking of other roles and in a preliminary analysis of the data for ergative *-ne* marking, we observed that the learners perform much better for ergative case marking (notice how Learner d, who clearly has not acquired Hindi DOM, uses ergative case marking in the example above). So transfer (in this case, operationalised as the absence of a case marking system in the L1 influencing the use of case marking in the LX) alone does not suffice to explain the high omission rates in the dataset of our study. A highly controlled experiment would be required where a group with a similar profile as our participant group is contrasted with a group that might be able to positively transfer knowledge of DOM, e.g. Spanish-speaking students of Hindi. But even then language learning interacts with a lot of different variables, and it is important to distinguish L1 transfer from processes that are inherent to language and the development of the interlanguage. We also observed frequent use of the Hindi numeral *ek* in the data, which in L1 Hindi mainly acts as a numeral but seems to be applied by the learners as a kind of indefinite article. A new referent is often introduced accompanied by the numeral *ek*. Whereas in Dutch the numeral *een* and the indefinite article *een* are homographs, Hindi does not have an article system and instead uses optional local markers (numerals, demonstratives and case marking) and word order to mark definiteness in the discourse (Dayal 2018). Recent research on the use of *ek* by native speakers of Hindi (Shukla et al. 2022), however, suggests that the numeral is grammaticalizing into an indefinite article. Rather than introducing the referent with a bare noun (which usually gets an indefinite reading in Hindi), the new referent might then be introduced with *ek*. In their study, Shukla et al. (2022) test the Optionality Hypothesis which claims that obligatory markers are acquired before optional markers. They do this through a series of experiments in which they compare L1 speakers of Hindi to speakers of English, Spanish and Madinka, languages with either obligatory newness markers or no newness markers at all. The language of child learners of Hindi differed from the language of adult speakers of Hindi, significantly more so than with the other languages, and the authors attribute this to the inconsistency (and hence optionality) of the input. The idea of optionality being difficult to acquire and resulting in

variable language use is also relevant to our study. We do not disregard L1 influence (in our study, influence of Dutch) as a possible explanation for the high omission rates in our study, rather, we think that L1 influence may enhance the overextension of the optionality of the use of DOM *-ko* in Hindi as a foreign language. Whereas most of the studies in our overview expected to find errors of commission (overgeneralisations), the learners of our study use *-ko* in different contexts and for different functions, but seem to overextend the optionality of the use of *-ko* for specific, inanimate DOs to the obligatory (and in some cases optional) use of *-ko* for specific and non-specific animate DOs, resulting in extended null-marking of the DO. This could also explain why *-ko* marking of the IO and the Experiencer develops towards higher accuracy rates, since there is no optionality for these functions. The fact that the use of *-ko* also remains relatively low for the Goal, where *-ko* is often dropped, confirms that the optionality of the use of *-ko* may be what makes the acquisition of this marker particularly difficult for the learners.

Another reason why we assume additional explanations apart from L1 influence or transfer for the high omission rates is that research on the acquisition of DO *-ko* marking by child learners of Hindi also reports errors of omission, which suggests there may be some general linguistic tendencies related to the acquisition of Hindi DOM. The question is hence not only why the learners barely expand the use of the case marker, but also why they restrict the use of the case marker to a limited number of contexts. In this regard, the usage-based approach that is put forward by Narasimhan (2005; 2020) can offer an explanation. In her longitudinal studies, Narasimhan (2005; 2020) investigates two possibilities: either the child learners rely on the notion ‘patient of transitive action’ whereby she expects the learners to overgeneralize *-ko* marking, or child learners may focus on distributional patterns in the input and thus only mark a subset of DOs. Rather than relying on a broad notion of patient and overextending the use of DOM *-ko*, Narasimhan observes a restricted use of DOM *-ko* and assumes that the learners expand their use of the case marker according to these distributional patterns. To account for the acquisition of *-ko* in our data, we would then expect the *-ko* marker to emerge with a specific lexical type of DO, or with a specific predicate lemma. Again, with 120 *-ko* marked utterances on a total of 1526 utterances we are not able to make hard claims. However, throughout

the four data points we do see that there are a few lemmas that receive the *-ko* marker more often than others. Our item-based analysis indeed showed that *-ko* marking was restricted to a particular set of noun and verb lemmas. Ten noun lemmas (of the 150 lemmas) accounted for 84% of all *-ko* uses and ten verb lemmas (of the 118 verb lemmas) accounted for 71% of all *-ko* uses. Regarding input, we already mentioned above that the instruction on DOM *-ko* marking in English teaching materials on Hindi and in the teaching in class is rather limited (especially when compared to the explicit instruction on ergative *-ne* marking and Recipient/Experiencer *-ko* marking). Despite this limitation, our data show that the learners have picked up on the marker, albeit to a very limited extent. This suggests that the learners may be acquiring the marker through *implicit learning* and have linked *-ko* marking to a subset of contexts that they probably acquired, implicitly, from their learning materials or out-of-school activities such as watching Hindi movies or listening to Hindi music. If we apply the notion of ‘patient of transitive action’, and taking the combination of noun type and verb type into account that most frequently appeared in an utterance with *-ko* marking (see Appendix), it could very well be the case that DO *-ko* marking is associated with the extent to which the patient is affected by an action. Via the item-based analysis we observed that the following lemmas receive the *-ko* marker more often than others: *ādmī* ‘man’, *cūmnā* ‘to kiss’ and *pīṭnā/mārṇā* ‘to hit/beat’. In the case of the combination ‘to hit a/the man’ or ‘to kiss a/the man’ (the contexts that most often received DO *-ko* marking in this study), the level of affectedness of the patient is high. Another factor that is important to observe is that sentences such as these have two human animate arguments. It is as if the use of *-ko* only occurs under very strict circumstances. Narasimhan (2020) too observes in her data that *-ko* is initially only used with animate patients and with pronouns. She proposes an interaction of the clines as described by Aissen (2003), i.e. the animacy cline and the definiteness cline:

Human > Animate > Inanimate

Pronoun > Name > Definite > Indefinite Specific > Non-specific

Narasimhan suggests that the most favorable contexts for DO *-ko* marking are those where the two clines converge. This could explain why we only find animate nouns and pronouns in our item-based analysis. Recall that the verb *mārnā* ‘to kill, to beat’ and the verb *bulānā* ‘to call someone’ are verbs that according to Mohanan (1994) always require *-ko* marking⁶. Both turned out to be among the top ten verbs that were realised with an overtly *-ko* marked DO in the results of our item-based analysis. Example (3) and (4) show that the verb *cūmnā* is realised with a *-ko* marked DO in Time 1 and Time 4 by the same learner:

(3) *Larkī.∅ larke=ko cūm rahī thī* [1a 159]
Girl.NOM boy=OBJ to kiss.PROG.PST
‘The girl was kissing the boy.’

(4) *siṃh=ne ādmī=ko cumā* [1a 4825]
Lion=ERG man=OBJ kiss. PERF.PRS
‘The lion kissed the man.’

In this context, we can also refer to the transitivity parameters as proposed by Hopper & Thompson (1980), who mention the importance of the number of participants, aspect, agency of the subject, the affectedness of the direct object and the individuation of the direct object. Another important factor to consider is the determining function of DOM. In contexts where the sentence contains two animate arguments, *-ko* also serves as a determiner of which argument has the subject function and which argument has the object function.

This finding brings us back to the complex form-meaning mappings that all of the authors writing about DOM in Hindi touch upon. The data seem to confirm that the main difficulty for the learners lies in learning all of the contexts where the use of *-ko* is required. In contexts where there is no optional/differential case marking, which is the case for the marking of the IO (dative) and for the Experiencer (dative subject), the learners seem to quickly grasp that *-ko* is required, even though omissions do occur – but omissions and mistakes are arguably inevitable when one is learning a language and an inherent part of the development

6 See appendix for an overview of the verbs.

of the interlanguage. Here, the hypothesis by Baten & Ponnet (2023) provides an interesting explanation. The authors predict the acquisition of Hindi case marking through developmental stages and assume that the learners acquire DOM *-ko* marking via semantic mapping of the patient. By semantic mapping they mean that the use of *-ko* is constrained to one semantic feature (as opposed to functionally applied according to the different constraints), in other words, they map the use of *-ko* onto one semantic feature. As the learners do not yet have access to all the information that they need in order to apply *-ko* marking correctly, and following Verbeke & Ponnet's (2022) analysis of DOM, they assume that the use of *-ko* will be restricted to specificity. Moreover, because of the multifunctionality of *-ko*, they assume that the learners will initially associate *-ko* with only one function, the dative marking of the IO, following the one-form-one-function principle. The acquisition of DOM *-ko* will then be triggered by the emergence of *-ko* as a Goal. These predictions have been tested in a cross-sectional dataset of 15 Hindi learners by Ponnet & De Cuypere (2023a). Although we cannot directly transfer the results from a cross-sectional study to a longitudinal study like the present one, they too observed a clear preference for the marking of specific, human animate DOs in their data. Rather than finding semantic mapping to specific DOs only, specific human animate DOs seem to be the most favorable context for DO *-ko* marking – and this is confirmed by our study. According to certain analyses of DOM in Hindi, particularly the one applied by López (2012) and Montrul et al. (2012; 2015; 2019) it is also the only context where the use of *-ko* is obligatory. In this sense, the prediction that *-ko* is initially mapped onto specific DOs is not wrong, but it should be fine-grained to specific human animate DOs. The prediction that *-ko* is associated with only one function in the initial stages of acquisition is not supported by our data (which may also be due to the fact that the learners already had one full semester of Hindi instruction). On the contrary, we found that *-ko* emerges early on as a marker of all of the functions that we mentioned, but that the trajectories from the moment of emergence onwards differ greatly. Instead of mapping *-ko* to only one function, it may be that the different functions of *-ko* are competing with each other. In Time 1 and Time 2, we observed omissions for DOM *-ko*, but also for Experiencer *-ko* and *-ko* as a marker of the Goal. Consider example (5):

- (5) *aur ādmī.ø* (..) *kutte=ko* (.) *kutt-e* (...) [t4 e 4341]
 And man.NOM dog=OBJ dog-OBL
ādmī.ø *kuttā* *aurat=ko* (...) *ehm detā hai*
 man.NOM dog.NOM women=OBJ ehm to give.PRS
 ‘And the man (..) the dog (.) (the) dog (...) the man gives the dog to
 the woman.’

Although we cannot be entirely sure about what was intended with the first instantiation of *kuttā* ‘dog’, this example is an indication of how the two functions of *-ko* can compete with each other. Ponnet & De Cuypere (2023a) performed detailed individual learner analyses and observed that a sentence with an overt IO *-ko* sometimes created the linguistic environment for *-ko* marking, and assumed this could trigger a DO *-ko* to be overtly realised. Baten (2013) suggests that the acquisition of case marking occurs via a series of oppositions. Our data suggests that the learners expand the use of *-ko* through a series of oppositions and reach high accuracy depending on whether *-ko* marking for the function is obligatory or optional, and whether the semantic and syntactic features regulating the use of *-ko* coincide.

7. Conclusion

This study contributes to the debate on the acquisition of DOM in Hindi in the following ways. First, we demonstrate that the acquisition of DOM in Hindi emerges when the features animacy and specificity coincide. We observed an interaction with verb semantics, the relative animacy of the DO, and the level of affectedness of the DO (in relation to the transitivity of the verb). Second, we tested different explanations provided in the literature for the acquisition of DOM in Hindi and verified findings of previous cross-sectional studies on the acquisition of HFL. Our longitudinal study confirms high rates of omission for all types of DO, also for the obligatory contexts. Despite a significant increase of the use of *-ko* between Time 3 and Time 4, even the better-performing learners in our study tend to prefer null-marking. Our findings suggest that the difficulty of the syntactic and semantic constraints drives the learners to initially overextend the optionality of *-ko* (which may be enhanced by L1 transfer) before they gradually

start using *-ko* according to its constraints (which seems to emerge via item-based learning).

Our findings also raise further questions regarding:

- 1) the acquisition of the Experiencer, which showed a sudden significant increase towards the end of our study. This role covers different constructions and a more in-depth study that covers these constructions is needed;
- 2) the interaction of the acquisition of Hindi DOM with ergative case marking. At present, we are performing such a study and have found that learners acquire ergative case marking in Hindi faster than DOM. More in-depth analyses are needed to understand why the acquisition of ergative marking seems to be less problematic for HFL learners than the acquisition of DOM;
- 3) The issue of input also requires attention, as our study has shown that learner materials insufficiently cover DOM. It would be valuable to investigate whether different ways of explicit instruction can improve the use of this case marker, which currently seems to be acquired in a more implicit way and whose intricacies are not sufficiently addressed in the learner materials that are available.

Aaricia Ponnet
Department of General Linguistics
Blandijnberg 2, 9000 Gent - Belgium
Aaricia.Ponnet@UGent.be

Ludovic De Cuypere
Department of Linguistics – Ghent University
Blandijnberg 2, 9000 Gent – Belgium
Brussels Centre for Language Studies
Pleinlaan 2, 1050 Brussels – Belgium
Ludovic.DeCuypere@UGent.be

References

- Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural language & linguistic theory*. 435-483. (doi:10.1023/A:1024109008573)
- Baten, Kristof. 2013. *The Acquisition of the German Case System by Foreign Language Learners*. Amsterdam: John Benjamins Publishing. (doi: 10.1075/palart.2)
- Baten, Kristof & Ponnet, Aaricia. 2023. Extending PT to split ergative marking and differential object marking. In Kawaguchi, Satomi & Di Biase, Bruno & Yamaguchi, Yumiko (eds.), *Processability and Language Acquisition in the Asia-Pacific Region 9*, 91-114. Amsterdam: John Benjamins Publishing. (doi:10.1075/palart.9.04bat)
- Baten, Kristof & Verbeke, Saartje. 2015. The acquisition of the ergative case in Hindi as a foreign language. In Baten, Kristof & Buyl, Aafke & Lochman, Katja & Van Herreweghe, Mieke (eds.), *Theoretical and methodological developments in Processability Theory*, 71-104. Amsterdam: John Benjamins Publishing. (doi: 10.1075/palart.4.04bat)
- Bates, Douglas & Maechler, Martin & Bolker, Ben & Walker, Steve. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1). 1-48. (doi:10.18637/jss.v067.i01.)
- Bautista-Maldonado, Salvador & Montrul, Silvina. 2019. An experimental investigation of differential object marking in Mexican Spanish. *Spanish in Context* 16(1). 22- 50. (doi: doi.org/10.1075/sic.00025.bau)
- Bhatt, Rajesh. 2005. Long distance agreement in Hindi-Urdu. *Natural Language & Linguistic Theory* 23(4). 757-807. (doi:10.1007/sl 1049-004-4136-0)
- Butt, Miriam & King, Tracy Holloway. 2004. The status of case. In Dayal, Veneeta & Mahajan, Anoop (eds.), *Clause structure in South Asian languages*, 153-198. Dordrecht: Springer. (doi:10.1007/978-1-4020-2719-2_6)
- Croft, William. 1988. Agreement vs. Case Marking and Direct Objects. In Barlow, Michael & Ferguson, Charles (eds.), *Agreement in Natural Language: Approaches, Theories, Descriptions*, 159-179. Stanford: Center for the Study of Language and Information (CSLI). (doi:10.5281/zenodo.4659627)
- Dayal, Veneeta. 2018. (In)definiteness without articles: diagnosis, analysis, implications. In Sharma, Ghanshyam & Bhatt, Rajesh (eds.), *Trends in Hindi linguistics*, 1-26. Boston: De Gruyter Mouton. (doi:10.1515/9783110610796-001)
- De Hoop, Helen & Narasimhan, Bhuvana. 2005. Differential case-marking in Hindi. In Amberber, Mengistu & De Hoop, Helen (eds.), *Competition and variation in natural languages*, 321-345. Amsterdam: Elsevier.

- Dulieu, Jean. 1972. *Paulus de Boskabouter. De baard van Pieter*. Bussum: Van Holkema & Warendorf.
- Dulieu, Jean. 1974. *Paulus de Boskabouter 7: Ali Baba en de Rover*. Bussum: Van Holkema & Warendorf.
- Everaert, Christine. 2004. *Basisgrammatica van het Hindi*. Gent: Academia Press.
- García, Marco García. 2018. Nominal and verbal parameters in the diachrony of differential object marking in Spanish. In: Seržant, Ilja A. & Witzlack-Makarevich, Alena (eds.), *Diachrony of differential argument marking*, 209-242. Berlin: Language Science Press. (doi:10.5281/zenodo.1228257)
- Harrell, Jr Frank. 2022. Hmisc: Harrell Miscellaneous. R package version 4.7-2, <<https://CRAN.R-project.org/package=Hmisc>>.
- Haspelmath, Martin. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In Tomasello, Michael, *The new psychology of language*, 217-248. Lawrence Erlbaum. (doi: 10.5281/zenodo.831410)
- Hopper, Paul J. & Thompson, Sandra A. 1980. Transitivity in grammar and discourse. *Language* 56(2). 251-299. (doi:10.1353/lan.1980.0017)
- Koster, Jan. 1975. Dutch as an SOV language. *Linguistic analysis* 1(2). 111-136. (doi: 10.5117/IVN2012.0.RUIG)
- López, Luis. 2012. *Indefinite objects: Scrambling, choice functions, and differential marking*. MIT press. (doi:10.7551/mitpress/9165.001.0001)
- Lüdecke, Daniel. 2018. ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *Journal of Open Source Software* 3(26). 772. (doi:10.21105/joss.00772)
- Lüdecke, Daniel. 2022. sjPlot: Data Visualization for Statistics in Social Science. R package version 2.8.12, <<https://CRAN.R-project.org/package=sjPlot>>.
- Klein, Udo & De Swart, Peter. 2011. Case and referential properties. *Lingua* 121(1). 3-19. (doi: 10.1016/J.LINGUA.2010.07.005)
- Malchukov, Andrei L. 2008. Animacy and asymmetries in differential case marking. *Lingua* 118(2). 203-221. (doi:10.1016/j.lingua.2007.02.005)
- Mardale, Alexandru & Montrul, Silvina. 2020. *The Acquisition of Differential Object Marking*. Amsterdam: John Benjamins Publishing Company. (doi:10.1075/tilar.26)
- Mohanan, Tara. 1994. *Argument Structure in Hindi*. Stanford: Center for the Study of Language.
- Montrul, Silvina & Bhatt, Rakesh Mohan & Bhatia, Archana. 2012. Erosion of case and agreement in Hindi heritage speakers. *Linguistic Approaches to Bilingualism* 2(2). 141-176. (doi:10.1075/lab.2.2.02mon)

- Montrul, Silvina & Bhatt, Rakesh Mohan, & Girju, Roxana. 2015. Differential object marking in Spanish, Hindi, and Romanian as heritage languages. *Language* 91(3). 564-610. (doi: 10.1353/lan.2015.0035)
- Montrul, Silvina & Gürel, Ayşe. 2015. The acquisition of Differential Object Marking in Spanish by Turkish speakers. In Judy, Tiffany & Perpiñán, Silvia S. (eds.), *The Acquisition of Spanish by speakers of less commonly studies languages*, 281-308. Amsterdam: John Benjamins. (doi: 10.1075/ihll.3.11mon)
- Montrul, Silvina & Bhatia, Archana & Bhatt, Rakesh, & Puri, Vandana. 2019. Case marking in Hindi as the weaker language. *Frontiers in psychology* 10. 1-19. (doi:10.3389/fpsyg.2019.00461)
- Montrul, Silvina. 2019. The acquisition of Differential Object Marking in Spanish by Romanian speakers. *Revista Española de Lingüística Aplicada* 32(1). 185-219. (doi:10.1075/resla.17040.mon)
- Narasimhan, Bhuvana. 2005. Splitting the notion of ‘agent’: case-marking in early child Hindi. *Journal of Child Language* 32(4). 787-803. (doi: 10.1017/s0305000905007117)
- Narasimhan, Bhuvana. 2020. *The acquisition of differential case-marking: Evidence from Hindi* [Conference keynote presentation]. SAFAL 2020, online via Zoom.
- Pallotti, Gabriele. 2007. An operational definition of the emergence criterion. *Applied Linguistics* 3(28). 361-382. (doi: 10.1093/applin/amm018)
- Pienemann, Manfred. 1998. *Language processing and second language development*. Amsterdam: John Benjamins. (doi: 10.1075/sibil.15)
- Ponnet, Aaricia & Baten, Kristof & Verbeke, Saartje. 2016. The acquisition of differential object marking in Hindi as a foreign language. *Dutch Journal of Applied Linguistics* 5(2). 101-125. (doi: 10.1075/dujal.5.2.01bat)
- Ponnet, Aaricia & De Cuypere, Ludovic. 2023a. The acquisition of Hindi split-ergativity and differential object marking by Dutch L1 speakers: systematicity and variation. *Language Acquisition*. 1-32. (doi:10.1080/10489223.2023.2195390)
- Ponnet, Aaricia & De Cuypere, Ludovic. 2023b. Replication Data for: The acquisition of Differential Object Marking: a longitudinal study on L1 Dutch learners of Hindi as a foreign language. *DataverseNO, VI*. (doi:10.18710/Q1AVUU)
- Sharma, Naresh. 2018. *Hindi tutor: advanced beginner to upper intermediate course. Grammar and vocabulary workbook*. London: Teach Yourself.
- Shukla, Vishakha & Long, Madeleine & Rubio-Fernandez, Paula. 2022. Children’s acquisition of new/given markers in English, Hindi, Mandinka and Spanish: Exploring the effect of optionality during grammaticalization. *Glossa Psycholinguistics* 1(1). 1-42. (doi:10.5070/G6011120)

- Sloetjes, Han & Wittenburg, Peter. 2008. Annotation by category - ELAN and ISO DCR. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).
- Tippets, Ian. 2011. Differential object marking: Quantitative evidence for underlying hierarchical constraints across Spanish dialects. In *Selected proceedings of the 13th Hispanic linguistics symposium*. 107-117. Cascadilla Proceedings Project.
- Verbeke, Saartje & Ponnet, Aaricia. 2022. Animacy, Specificity, and Verb Semantics: What Drives Differential Object Marking in Hindi?. In Abbi, Anvita & Oranskaia, Tatiana (eds.), *The Heart of Change - Issues on Variation in Hindi*, 133-154. Heidelberg: Heidelberg Asian Studies Publishing. (doi:10.11588/hasp.919.c12621)
- Wickham, Hadley & François, Romain & Henry, Lionel & Müller, Kirill. 2022. dplyr: A Grammar of Data Manipulation. R package version 1.0.10, <<https://CRAN.R-project.org/package=dplyr>>.
- Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.
- Zwart, C. Jan-Wouter. 1997. *Morphosyntax of verb movement: A minimalist approach to the syntax of Dutch*. Dordrecht: Kluwer.

Appendix

1. *DOM in Hindi*

DOM is a common phenomenon whereby the DO is morphologically marked depending on features such as animacy, specificity, definiteness, and/or referentiality (Malchukov 2008). The interplay of these features result in an interaction between the grammatical function DO and the semantic role patient. In many languages, DOM marks patients possessing features that renders them agent-like, thus discriminating between agent and patient (Tippets 2011). Important in this regard is also the ‘relative animacy’ between the agent and the patient. Animate referents (and especially human animate referents) are more likely to appear in the role of agent. When a sentence contains two animate arguments, the DO marker discriminates between the agent and the patient. The role of animacy in DOM can then be explained by means of the animacy hierarchy (Aissen 2003: 442; based on Croft 1988):

Human > Animate > Inanimate

In many languages that have DOM, arguments that rank higher on the animacy scale are more likely to be marked. Aissen described a scale for definiteness as well, and related this to DOM:

Pronoun > Name > Definite > Indefinite Specific > Non-specific

Here, too, DOs that are ranked higher are more likely to be marked. DOM in Hindi is claimed to be one of the best described examples of DOM that combines both animacy and definiteness/specificity (Malchukov 2008). However, with respect to the latter factors, Verbeke & Ponnet (2022) observed that neither definiteness nor specificity are clearly defined in the literature and that there exists no overall agreement on whether Hindi DOM is driven by definiteness or by specificity. Both terms arguably overlap to a certain extent.

A problem with specificity is that the interpretation is actually context-dependent. Consider in this regard example (6), which involves an animate DO *bacce* (‘child’), and which is marked by *-ko*:

- (6) *ilā=ne bacce=ko / *bacc-ā.ø uṭhāyā* (source: Monahan 1994: 79)
 Ila=ERG child=OBJ lifted-PERF
 ‘Ila lifted a/the child’

The *-ko* marker is used because of the animacy of the DO. Whether the interpretation is specific or non-specific depends on the context. Specificity is only involved, however, when the DO has a generic reading, where *-ko* is dropped, such as in example (7):

- (7) *ravī.ø (ek) gāy.ø kharīdnā + cāhtā + hai*
 (source: Monahan 1994: 80)
 ravi.NOM (one) cow.NOM wishes to buy.PRS
 ‘Ravi wishes to buy a cow (with no particular cow in mind).’

Note that the DO in this example is non-human animate, and that the numeral *ek* is given in brackets. Moreover, Hindi DOM also appears to be related to the sentence verb. Mohanan (1994: 81) gives the following overview of verbs that either never or always have a *-ko* marked DO:

- Never *-ko*: *likhnā* ‘write,’ *paṛhnā* ‘read,’ *pīnā* ‘drink,’ *gānā* ‘sing’ and *banānā* ‘make’;
- Always *-ko*: *bulānā* ‘call’ and *mārṇā* ‘beat somebody,’ ‘kill’.

Verbeke & Ponnet (2022) compiled a corpus with these verbs (based on the EMILLE and COSH Corpus of Spoken Hindi) and found that, in those instances where the sentences did not align with Mohanan’s description, the exceptions for both verb types (never *-ko* vs. always *-ko*) could be explained by the feature specificity. This dovetails with Klein and deSwart (2011) who describe DOM in terms of *trigger* and *effect*. Whereas in Hindi the presence of the feature [+animate] *triggers* the *-ko* marking, the presence of *-ko* marking has a [+specific] reading as its *effect*. In these terms, one expects a *-ko* marked inanimate DO to render a specific reading (like in 8a), i.e. the specific reading is the *effect* of the presence of *-ko*, whereas the (non-)specific reading of an inanimate DO that is not *-ko* marked depends on the context and the intention of the speaker (see 8b and 8c). When a non-specific meaning is intended for an inanimate DO, one does not expect *-ko* at all (8c):

- (8) a. Inanimate/specific (with *ko*)
ilā=ne hār=ko uṭhāyā (source: Monahan 1994: 81)
 Ila=ERG necklace=OBJ lifted-PERF
 ‘Ila lifted *a/the necklace’
- b. Inanimate/specific (without *ko*)
ilā=ne hār.∅ uṭhāyā (source: Monahan 1994: 81)
 Ila=ERG necklace.NOM lifted-PERF
 ‘Ila lifted a/the necklace’
- c. Inanimate/non-specific
*ilā=ne hār.∅ / *hār=ko uṭhāyā* (source: Monahan 1994: 81)
 Ila=ERG necklace.NOM / *=OBJ lifted-PERF
 ‘Ila lifted a necklace’

Although many accounts on DOM in Hindi present its rules as being not overly complicated, and although most Hindi textbooks (e.g. Everaert 2004, Sharma 2018), for learners do not spend more than half a page on the use of this marker, Verbeke & Ponnet (2022) have shown that there is no consensus on which factor is stronger and how definiteness/specificity can be defined with regard to Hindi DOM. More research on this feature is definitely needed to account for the many exceptions to the rules presented above. Moreover, DOM in Hindi seems to interact with split ergativity (De Hoop & Narasimhan 2005), verbal agreement (Bhatt 2005) and other roles marked by the same marker *-ko* (cf. Butt & King 2004; Haspelmath 2003), as illustrated in table 11:

Table 11: *-ko* marking of other roles

<i>-ko</i> marking of other roles	
Recipient	<i>ko</i>
Experiencer	<i>ko</i>
Goal	∅ / <i>ko</i>

It can therefore be challenging for a foreign language learner to acquire the different semantic and syntactic features associated with Hindi case marking, as has been shown in previous research on the acquisition of Hindi DOM.

2. Annotation

Per participant and per Time we conducted an interview of about 30-45 minutes, which means we collected almost three hours of data per participant. The interviews were transcribed using the transcription tool ELAN (<https://archive.mpi.nl/tla/elan>) that was developed at the Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands. The utterances were then annotated for the following variables:

- Time: Point at which the data was elicited:
 - Time 1: Halfway the second semester of the first year
 - Time 2: End of the first semester of the second year
 - Time 3: Beginning of the second semester of the second year
 - Time 4: End of the second semester of the second year
- Learner: Unique identifier for every learner in the data (a, b, c, d, e)
 - DO_Item: The lemma of the Noun to which *-ko* is added. This Noun is the head of the Noun Phrase that functions as the Direct Object.
- DO_ko : whether *-ko* marking is used in the utterance by the learner. (“yes” vs. “no”)
- DO_Specificity: Specificity of the Direct Object. (“specific” vs. “non-specific”)
- DO_Animacy: Animacy of the Direct Object. We distinguished between three levels:
 - “human animate”
 - “inanimate”
 - “non-human animate”
- V_Item: The lemma of the sentence verb
- I_Item: The lemma of the Noun (Phrase) that can also take *-ko* marking. The Noun is the head of the Noun Phrase.
- I_ko: Whether the marker *ko* is present, but not for as a Differential Object Marker. (“yes” vs. “no”)
- I_SemRole: we distinguish three semantic roles of I_ko
 - “Goal”
 - “Experiencer”
 - “Recipient”

3. Item-based analysis

In total, we elicited 150 different noun lemmas for the DO and 118 verb lemmas that were realised with a DO. In table 12 below, we have given an overview of the ten DO lemmas that were most marked with *-ko*:

Table 12: Top ten Nouns occurring with *-ko* marking.

	Noun	DO-∅	DO-ko
1	<i>ādmī</i> ‘man’	88 (66%)	45 (34%)
2	<i>vah</i> ‘he/she’	3 (20%)	12 (80%)
3	<i>jānvar</i> ‘animal’	82 (85%)	11 (15%)
4	<i>cūhā</i> ‘mouse’	62 (88%)	11 (12%)
5	<i>baunā</i> ‘gnome’	16 (70%)	7 (30%)
6	<i>bandar</i> ‘monkey’	16 (80%)	4 (20%)
7	<i>laṛkī</i> ‘girl’	13 (81%)	3 (19%)
8	<i>siṃh</i> ‘lion’	9 (62%)	3 (38%)
9	<i>grāmvāsī</i> ‘villager’	5 (75%)	3 (25%)
10	<i>gaimṛā</i> ‘rhinoceros’	5 (71%)	2 (29%)

Taken together, these 10 DO Noun types account for 101 (84%) of all *-ko* uses (recall that there are 120 observations of DO-ko). All of these noun types are animate, of which the noun *ādmī* ‘man’ counts for most of the *-ko* marked DOs. To assess whether the semantics of the verb influenced DOM *-ko* marking, we also analysed which verbs occurred most with a *-ko* marked DO, as illustrated in table 13:

Table 13: Top ten Verbs occurring with *-ko* marking.

	Verb	DO -∅	DO -ko
1	<i>dekhnā</i> ‘to see/look/ watch’	398 (95%)	20 (5%)
2	<i>cūmnā</i> ‘to kiss’	12 (41%)	17 (59%)
3	<i>pīṭnā</i> ‘to hit/beat (some- one)’	15 (62%)	9 (38%)
4	<i>bulānā</i> ‘to call (some- one)’	18 (72%)	7 (28%)
5	<i>sunnā</i> ‘to hear’	77 (92%)	7 (8%)
6	<i>khānā</i> ‘to eat’	70 (92%)	6 (8%)
7	<i>lenā</i> ‘to take’	122 (95%)	6 (5%)
8	<i>mārnā</i> ‘to kill/to beat’ (someone)	1 (17%)	5 (83%)
9	<i>dikhānā</i> ‘to show’	57 (93%)	4 (7%)
10	<i>paḡarṇā</i> ‘to grab/take’	33 (89%)	4 (11%)

Taken together, the 10 verbs account for 85 (71%) of all *-ko* uses. Here, the uses of *-ko* are more spread, with the verbs *dekhnā* ‘to see’, *cūmnā* ‘to kiss’ and *pīṭnā* ‘to hit’ (note that *mārnā* ‘to beat someone’ has a similar meaning) occurring mostly with a *-ko* marked DO.