A corpus-based analysis of dative alternation in Indian English

LUDOVIC DE CUYPERE and SAARTJE VERBEKE

ABSTRACT: The dative alternation refers to the alternation between two constructions that denote some type of transfer: the double object construction (*I give my sister a book*) vs. the to-dative construction (*I give a book to my sister*). We examined the motivations behind the dative alternation in Indian English. A corpus study was performed based on a sample of $N = 943$ sentences that were drawn from the *Kolhapur* corpus. Using a mixed-effects logistic regression analysis, we evaluated the effect of 14 predictors that are known to influence the dative alternation in other macro-regional varieties of English. Three predictors were found to be significant: the verb (modeled as a random intercept), the pronominality of the Recipient and the difference in length between the Recipient and the Theme. Our results further corroborate earlier findings that the to-dative construction is more frequently used in Indian English than in other varieties. We argue that the latter tendency may be associated with a transfer from Hindi.

Keywords: dative alternation, Indian English, Ditransitivity, to-dative construction, mixed-effects logistic regression, Hindi, transfer.

Background

The dative alternation is defined as the alternation between two constructions that denote some type of transfer: the double object construction or "ditransitive construction" (DOC), which is exemplified in (1) and the prepositional dative or 'to-dative construction', which is exemplified

in (2). Semantically, both constructions conglomerate three semantic roles, which together represent a transfer setting: Agent (‘doer’), Theme (‘entity’) and Recipient (‘receiver’).

(1) I give [my sister]$_{\text{RECIPIENT}}$ [a book]$_{\text{THEME}}$.

(2) I give [the book]$_{\text{THEME}}$ [to my sister]$_{\text{RECIPIENT}}$.

As the examples illustrate, the alternation actually involves two differences: a change in constituent order and the use of the preposition *to*. A DOC places the Recipient before the Theme and is prepositionless. In contrast, the to-dative construction adds *to* as a Recipient marker and places the Theme before the Recipient.[1]

It has often been argued that the English dative alternation can be captured in terms of verb semantics alone (e.g., Green 1974; Oehrle 1976; Pinker 1989). Thus, a *deterministic* relationship was assumed between the meaning of a verb and the co-occurring ditransitive construction. This hypothesis was corroborated by the intuition of several scholars that certain verb classes can felicitously be used only in one specific ditransitive object order. For instance, verbs that involve the continuous impartation of force (e.g., *carry*, *pull*, *push*, *lift*) were considered to be ungrammatical with the DOC, as illustrated in (3) (examples obtained from Bresnan, Cueni, Nikitina and Baayen 2007: 71):

(3) *I carried/pulled/pushed/schlepped/lifted/lowered/hauled John$_{\text{RECIPIENT}}$ the box$_{\text{THEME}}$.

In this deterministic view, verbs that involve the continuous imparting of force prompt the use of the prepositional ditransitive construction because this verb class would semantically be

associated with a movement toward a goal rather than with a transfer of possession event. However, by scanning vast corpora of electronic language data and by browsing the internet, Bresnan et al. (2007: 73) found numerous cases in which this alleged ungrammaticality did not apply, such as in (4), in which the recipient is expressed as a pronominal object:

(4) As player A pushed him$_{\text{RECIPIENT}}$ the chips$_{\text{THEME}}$, all hell broke loose at the table.

Bresnan, Cueni, Nikitina, and Baayen (2007) found that although there exist strong preferences for certain verbs to occur with one particular ditransitive construction, other factors also bear an effect on a speaker's choice of construction. Therefore, multiple factors must be considered *simultaneously* to provide a full account of the dative alternation.

To date, a wide variety of factors that are involved in the dative alternation have been uncovered, including verb semantics (Levin, 1993; Lapata, 1999; Gries, 2005; Bresnan, Cueni, Nikitina, and Baayen 2007); the discourse status of the theme/recipient, i.e., whether the object introduces a new referent or refers to a given referent (Halliday 1970; Erteschik-Shir 1979; Smyth, Hogan and Prideaux 1979; Givón 1984; Thompson and Koide 1987; Thompson 1995; Bresnan, Cueni, Nikitina, and Baayen 2007; Ozón 2009); the pronominality and definiteness of the theme/recipient (Ransom 1979; Bresnan, Cueni, Nikitina, and Baayen 2007); the animacy and person of the recipient (Bresnan and Nikitina 2007); and the weight (in terms of either length or syntactic complexity) of the theme/recipient (Bock and Irwin 1980; Bock, Loebell and Morey 1992; Hawkins 1994; Collins 1995; Thompson 1995; Arnold, Losongco, Wasow and Ginstrom 2000; Prat-Sala and Branigan 2000; Wasow 2002; Snyder 2003; Ozón 2009). Based on these factors, Bresnan et al. (2007) successfully fitted a mixed-effects logistic regression model that

was able to predict 94% of the actual choices of dative constructions in their dataset of natural spontaneous conversations obtained from the Switchboard collection of recorded telephone conversations. Additional corpus studies (Theijssen 2008, 2009; Kendall, Bresnan and Van Herk, 2011; de Marneffe, Grimm, Arnon, Kirby and Bresnan 2012) and psycholinguistic experiments (Bresnan and Ford 2010; Theijssen, Bresnan, Ford and Boves, subm.) have repeatedly corroborated these findings.

In this article, we report on a corpus-based study of the dative alternation in Indian English (IndE). Our dataset consists of a sample of N = 943 observations obtained from the *Kolhapur* corpus of IndE. Our overall hypothesis is that the same factors that motivate the use of the dative alternation in other varieties of English will also be involved in the dative alternation in IndE. However, the existence of differences between the norms of different varieties of a language (e.g., regional and social differences) is well known. With regard to the dative alternation, a difference in the effect of end weight between various macro-regional varieties has been found. Our study enables us to discover how the motivations behind the dative alternation in IndE differ from those behind the dative alternation in other English variaties under review.

To our knowledge, there is no published research on the IndE dative alternation. However, there is a significant amount of research on ditransitive verb complementation (Nihalani, Tongue and Hosali 1979; Olavarría de Ersson and Shaw 2003; Muhkerjee and Hoffman 2006;). There is evidence that IndE and BrE diverge from one another in the frequency and distribution of verb complementation patterns. Olivarría de Ersson and Shaw (2003: 153) found that the to-dative construction is more frequent in IndE than in BrE with the verbs *provide*, *entrust*, *furnish*, *supply* and *present*. Mukherjee and Hoffmann (2006) also found that the to-dative construction is more frequent with the verbs *give* and *send* in IndE than in BrE. Their

results further indicated that the DOC diffused more slowly among IndE verbs than BrE verbs; the verb *inform* is one instance of a "new ditransitive". Together, then, previous scholarship suggests that the to-dative construction is more pervasive in IndE than in other varieties of English.

One explanation claims that IndE has preserved the complementation pattern of 19th century BrE (Olavarría de Ersson and Shaw 2003: 158; Muhkerjee and Hoffman 2006: 166). However, this explanation is inconsistent with the finding that dative proportions have remained fairly stable throughout the history of English (Wolk, Bresnan, Rosenbach and Szrmecsányi, 2012). Moreover, we found no statistically significant difference between the results of our study and the relative proportion of to-dative constructions in 19th century BrE. A more likely explanation, which is also consistent with findings regarding other varieties of English, is that the dative alternation in IndE evolved away from BrE and developed its own specific preferences. We additionally discuss two complementary explanations for the frequent use of the to-dative construction in IndE. Both of these explanations are based on the assumption that the larger proportion of to-datives may be associated with a transfer from Hindi.

The next section outlines our methodology and hypotheses. Sections 4 and 5 subsequently present and discuss the results. The conclusions are drawn in section 6.

Methodology

Kolhapur corpus

Our data were retrieved from the Kolhapur corpus. This corpus is a one million-word sample of standard Indian English texts from diverse written sources (books, government documents, press

materials and periodicals) and was published in 1978 (strictly speaking, then, we examine a corpus of late 20th century Indian English).[2] ~~With respect to the authors~~

Data sample

We gathered our data by means of lexical searches based on 68 verbs that are known to occur with the dative alternation in British English (BrE) (cf. Wolk, Bresnan, Rosenbach, and Szrmecsányi 2012: 7): *give, tell, sell, pay, offer, cost, send, take, show, bring, charge, owe, loan, write, feed, mail, hand, cause, leave, wish, allow, read, deny, serve, assign, allot, lend, promise, quote, afford, award, flip, float, swap, grant, issue, extend, lease, allocate, deliver, resell, teach, assure, cede, deal, fine, guarantee, permit, accord, assess, bequeath, bet, carry, funnel, get, net, prepay, present, refuse, reimburse, repay, run, slip, submit, supply, tender, trade* and *vote*. Each verb was sought in all of its possible forms (e.g., *give, gives, gave, given and giving*). We found observations for 27 verbs and gathered a dataset of more than 2000 sentences. Following Theijssen (2008), we excluded all cases based on the seven criteria that are defined in the list (i to vii) below. Excluding all ineligible cases, we obtained a corpus sample of N = 943 observations.

i. DOCs with a *benefactive* indirect object were excluded because its prepositional paraphrase takes the preposition *for* rather than *to*, cf. (5):

(5) a. He bought him sweets. <Kolhapur.1030K12>

b. He bought sweets for him.

ii. To-prepositional constructions with a *locative to-prepositional phrase* (e.g., (6)) were excluded because the DOC paraphrase necessarily gives the location an animate reading:

(6) The district magistrate sent the lawyer himself to prison. <Kolhapur.0470F23>

We retained examples in which the to-prepositional phrase had a metonymical interpretation, such as in (7), in which the location's name actually refers to a group of people:

(7) The town must send relief supplies to the village. <Kolhapur.1070K25>

iii. *Non-canonical constituent orders* (i.e., orders that diverge from the canonical V-NP-NP and V-NP-PP realizations) were additionally excluded: passives, either with the Recipient as subject (8) or with the Theme as subject (9), and the order V-PP-NP (10):

(8) Capt. Cousteau's son was given his first dive at the age of four. <Kolhapur.0700E18>

(9) The decline of priests was given first priority in the list of subjects discussed. <Kolhapur.0470D17>

(10) He presented to Swayambhu a silken purse. <Kolhapur.0240K19>

iv. *Non-Noun Phrase Direct Object realizations*. Clausal objects typically occur in final sentence position and rarely alternate, such as in (11) and (12).

(11) A police official who told us that they had come to arrest us. <Kolhapur.0730G33>

(12) Tell him to get some medicine. <Kolhapur.0550K43>

v. Sentences containing *phrasal verbs*, which were defined as fixed combinations of a main verb with one or more prepositions, cf. *hand over* in (13):

(13) It is my great pleasure to hand over this office to my son. <Kolhapur.0400E30>

vi. *Set phrases*. These are fixed combinations of an empty verb, a noun and the preposition *to* (*to give birth to* and *give rise to*). These combinations always require a to-dative construction (Jespersen 1933), which generate idioms, as in (14):

(14) Saraswati gave birth to nine children. <Kolhapur.1010F36>

vii. Sentences in which the Theme/Recipient is composed of two or more *coordinated entities.* Such cases obscure the determination of the variables, as in (15):

(15) Here the British had given him a bungalow and the freedom to move about within the confines of the district. <Kolhapur.0920K04>

Both parts of the Theme are determined differently: *a bungalow* is concrete and indefinite, whereas *the freedom* is abstract and definite.

Predictor variables and associated hypotheses

All predictor variables and their respective operationalization are based on the work of Bresnan and Ford (2010). We did not annotate our corpus sample for discourse status (i.e., given, accessible and new) and structural parallelism (i.e., the occurrence of a DOC or to-dative construction in the previous discourse). Both have been found to influence the dative alternation. However, we found both discourse status and structural parallelism to be difficult to operationalize objectively for our written data. First, it is unclear how extensively one must search the text to determine whether the referent of an object is new or given. Second, it is not because the referent has not been mentioned in the previous lines of text that it was not already activated in the mind of the writer. The same reasoning applies to the structural parallelism predictor. Moreover, discourse status is also at least partly captured by pronominality, as pronominal objects are likely to refer to given discourse referents (pPronominality was found to be strongly correlated with givenness in the research of Theijssen, Bresnan, Ford and Boves subm.). In fact, as explained in Ssection 3, we believe that the result for pronominality is best understood in terms of discourse status.

VERB and SEMANTIC CLASS (VERB SENSE). The dative alternation is known to be associated with verbal semantics; for instance, *give* is more likely to appear in a DOC than *bring*. Semantic classes refer to the contextual interpretation of sentence verbs. For instance, *give* may be used in a concrete transfer sense (e.g., *give s.o. a book*) or in a more abstract sense (e.g., *give an explanation*). We distinguished between six semantic classes: the 'transfer' of possession of an object (e.g., *give a book to someone*); the 'future transfer' of objects (e.g., *owe money to someone*); the 'communication' of information (e.g., *tell a story to someone*); the 'prevention of possession' (e.g., *deny food to someone*); and 'abstract' (e.g., *allow someone a privilege*), which

contains nearly every other remaining sense. Although the verb *give* was frequently found in the sample, its usage as a verb of transfer was limited. Verb sense is defined as the combination of a verb and the semantic class with which it is used.

LENGTH DIFFERENCE between Recipient and Theme. Relative syntactic complexity was measured as "the natural logarithm of the recipient length minus the natural logarithm of the ~~recipient~~ theme length" (Bresnan and Ford 2010: 174). It is well known that constituents tend to appear in order of increasing length (Wasow 2002). The same effect is expected to be involved here.

ANIMACY of Theme and Recipient ('animate' vs. 'inanimate'). Humans, animals and organizations/institutions were conceived as animate. Thus, *the press* in (16) refers to a group of journalists and was regarded as 'animate'. Locations that referred to a group of people, such as *the press* in (16), were similarly regarded as 'animate'.

(16) He had given his reaction to the press on Mr. Unnikrishnan's comments in the house. <Kolhapur.0780A15>

CONCRETENESS of Theme ('concrete' vs. 'abstract'). A concrete entity is defined as an "inanimate object or substance that is perceivable with one of the five senses" (Bresnan and Ford 2010: 175). All other cases were regarded as abstract. This variable enables one to make a distinction between plants, which are considered inanimate, and more abstract words (e.g., *health, love, fear*). Concrete Themes are expected to occur with the to-dative construction.

DEFINITENESS of Theme and Recipient ('definite' vs. 'indefinite'). To determine this status, we used the substitution test of Garretson (2003): *There is/are__* (which can be felicitously be

used only with an indefinite referent). We used the context to determine unclear cases (e.g., *you*). A definite Theme is expected to prefer the to-dative construction, whereas a definite Recipient is expected to prefer the DOC.

PRONOMINALITY of Theme and Recipient ('pronominal' vs. 'nonpronominal'). In previous studies, pronominality was found to yield a strong effect on the dative alternation. All else being equal, we expect pronominal Themes or Recipients to occur before nominal Recipients or Themes.

PERSON of Theme and Recipient ('local' vs. 'non-local'). Local is defined as the first- or second-person singular or plural, and non-local refers to the third-person singular or plural. All else being equal, a local Theme/Recipient is expected to occur before a non-local Recipient/Theme.

NUMBER of Theme and Recipient ('singular' vs. 'plural'). Formal marking (i.e., the presence or absence of a plural marker) was used to determine the category of this variable. For unclear cases (e.g., *you*), we used the context. Singular Themes/Recipients are expected to precede plural ~~Recipient/~~Themes/Recipients.

Results

We collected a total of N = 943 ditransitive sentences: 526 (56%) sentences with an NP realization and 417 (44%) sentences with a PP realization. Table 1 compares this proportion with other known proportions of other English varieties.

Table 1: Relative frequencies of alternating constructions in the Indian English and British English samples

	Kolhapur (written)	Switchboard* (spoken)	WSJ† (written)	ICE-GB‡ (written)	ICE-GB‡ (spoken)
NP-NP	526 (56%)	1850 (79%)	561 (62%)	263 (72%)	399 (72%)
NP-PP	417 (44%)	499 (21%)	344 (38%)	102 (28%)	151 (28%)
Total	943 (100%)	2349 (100%)	905 (100%)	365 (100%)	550 (100%)

Sources: * Bresnan & Ford (2010)

† *Wall Street Journal*, Bresnan et al. (2007)

‡ Theijssen (2009)

All of the differences among the proportions of to-datives in the other samples are significant at the 1% significance level (the result for the proportion of the *Wall Street Journal* sample: Pearson's $\chi^2(1) = 7.35$, p-value < 0.007). Thus, our results confirm the observation that the to-dative construction is used more frequently in IndE than in other varieties under review.

Table 2 presents the bivariate results of our corpus study.

Table 2: bivariate results

	REALIZATION OF RECIPIENT	
	NP	PP
Semantic Class		

abstract	303 (51%)	289 (49%)
communication	96 (76%)	30 (24%)
future transfer	11 (58%)	8 (42%)
prevention	3 (60%)	2 (40%)
transfer	113 (56%)	88 (44%)
Animacy of Recipient		
animate	473 (62%)	291 (38%)
inanimate	48 (28%)	123 (72%)
Animacy of Theme		
animate	4 (29%)	10 (71%)
inanimate	517 (56%)	404 (44%)
Concreteness of Theme		
concrete	166 (57%)	126 (43%)
abstract	355 (55%)	288 (45%)
Definiteness of Recipient		
definite	497 (59%)	348 (41%)
indefinite	24 (27%)	66 (73%)
Definiteness of Theme		
definite	196 (46%)	232 (54%)
indefinite	325 (64%)	182 (36%)
Pronominality of Recipient		
pronominal	404 (89%)	51 (11%)
nominal	117 (24%)	363 (76%)

Pronominality of Theme		
pronominal	35 (39%)	54 (61%)
nominal	486 (57%)	360 (43%)
*Person of Recipient		
local	190 (93%)	15 (7%)
nonlocal	331 (45%)	399 (55%)
Number of Recipient		
singular	396 (59%)	277 (41%)
plural	125 (48%)	137 (52%)
Number of Theme		
singular	425 (56%)	338 (44%)
plural	96 (56%)	76 (44%)

*Person of Theme is not reported because all Themes were 'nonlocal'

The bivariate results are consistent with our general expectations: animate, definite, pronominal, local and singular Recipients are most often realized as NPs (i.e., in DOCs), with the highest percentages for pronominal and local Recipients. Conversely, inanimate, indefinite, nominal, nonlocal, and plural Recipients are most often realized as PPs, with the higher percentages for indefinite and nominal Recipients. The semantic class 'communication' appears to prefer an NP realization, whereas the opposite realization is preferred by the transfer class.

We used mixed-effects regression modeling to evaluate the significance of the predictors under examination (with the verb as a random intercept). Table 3 presents the results of the model that includes all of the variables under examination.[3]

Table 3: Mixed-effects logistic regression estimates for the full model: the PP realization is treated as the success.

	Coefficient (se)	*p*-value	Odds Ratio (95% CI)
Intercept	2.600 (3.816)	0.495	
Semantic Class			
communication	-1.294 (0.375)	0.078	
future transfer	-1.676 (1.182)	0.156	
prevention	-2.957 (3.248)	0.362	
transfer	-0.669 (0.642)	0.298	
Animacy Rec.			
inanimate	0.555 (0.389)	0.154	
Definiteness Rec.			
indefinite	1.042 (0.563)	0.064	
Pronominality Rec.			
pronominal	-2.363 (0.378)	4.13e-10	0.09 (0.04 to 0.20)
Person Rec			
nonlocal	0.629 (0.499)	0.207	
Number Rec.			
singular	-0.051 (0.341)	0.881	
Animacy Theme			
inanimate	-0.699 (2.920)	0.811	
Concreteness Theme			

concrete	0.749 (0.574)	0.192	
Definiteness Theme			
indefinite	-0.250 (0.304)	0.410	
Pronominality Theme			
pronominal	0.336 (0.574)	0.559	
Number Theme			
singular	-0.273 (2.456)	0.911	
plural	-0.651 (2.468)	0.792	
Length Difference	2.744 (0.248)	< 2e-16	15.55 (9.56 to 25.28)
Random intercept			
Verb (N=45)	1.695* (1.302)		

*Estimated variance and its standard error

Model diagnostics indicate an excellent goodness of fit: Somers' D_{xy} = 96%; index of concordance C = 98% (against a majority baseline of 56%). As shown in Figure 1, the mean expected proportion of successes is also correlated with the observed proportion of successes.

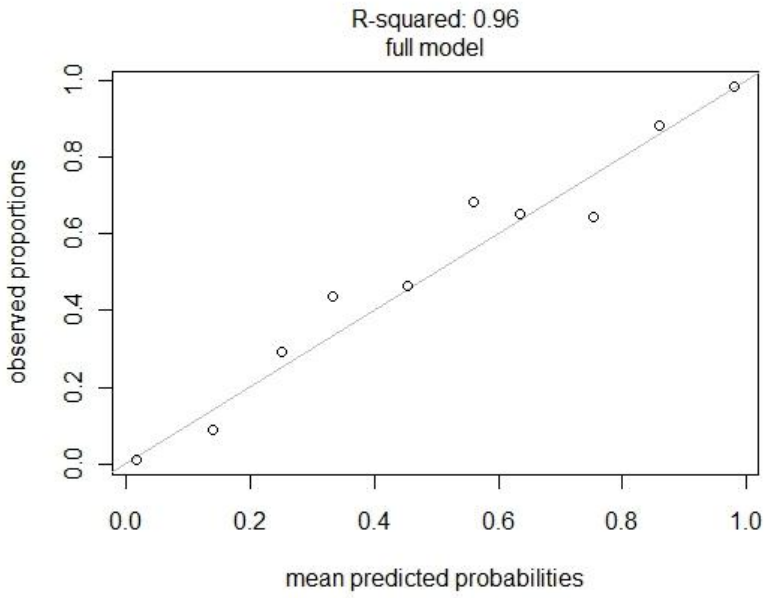

Figure 1: Corpus model fit between grouped observations and mean predicted probabilities.

Examining the significance of the predictor variables, however, we can observe that only two covariates were significant at a 5% significance level: the pronominality of the Recipient and differences in length (The definiteness of the Recipient and semantic class (communication) are significant at a 1% significance level). As expected, a pronominal Recipient was less likely to be expressed as a PP than a nominal Recipient. The odds ratio indicates that the odds of a pronominal Recipient being expressed as a PP are only 9% that of a nominal Recipient. We are 95% confident that these odds lie between 4% and 20%, an interval that is reasonably narrow.

The effect of the length difference was also as expected: the longer the Recipient is with respect to the Theme, the more likely it is that the Recipient will be expressed as a PP. However, the exact estimate of this predictor is more difficult to interpret because of the logarithmic transformation.

The variance of the Random factor verb is estimated at 1.695, which suggests that a considerable proportion of variability is associated with the verb as such. Note that the predictor semantic class was not found to be significant (except perhaps for the category communication, which was significant at a 1% significance level). This result implies that most of the variation is associated with specific verb semantics rather than with semantic classes; this observation was also made by Baayen (2008: 281) with regard to the Bresnan et al. (2007) dataset.

We additionally fitted a reduced model that retains only the significant variables. The results of this model are provided in Table 4:

Table 4: Mixed-effects logistic regression estimates for the minimally adequate model

	Coefficient (se)	*p*-value	Odds Ratio (95% CI)
Intercept	1.973 (0.366)	6.84e-8	
Pronominality Rec.			
pronominal	-2.541 (0.307)	<2e-10	0.08 (0.04 to 0.14)
Length Difference	2.870 (0.235)	< 2e-16	17.64 (11.13 to 27.95)
Random intercept			
Verb (N=45)	2.392* (1.547)		

*Estimated variance

The quality of this model was also good: Somers' D_{xy} = 95%; index of concordance C = 98% (majority baseline = 56%). However, the plot in Figure 2 suggests a lower goodness of fit for the reduced model than for the full model, particularly in the 0.2 to 0.4 interval. Nevertheless, an R^2

value of 80% indicates that the predicted probabilities of our reduced model are still close to the observed realizations.

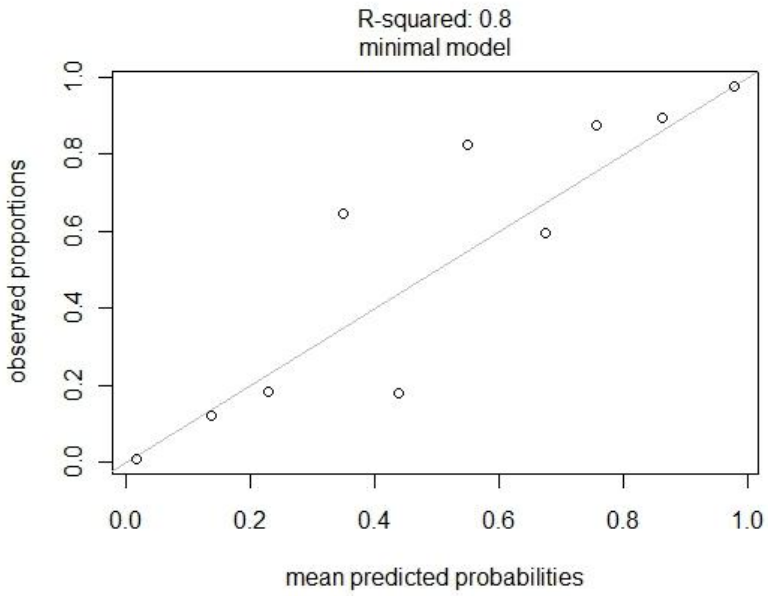

Figure 2: Corpus model fit between grouped observations and mean predicted probabilities.

The same predictors that were found to be significant in the full model are retained in the reduced model. A likelihood ratio test indicated no significant difference between the full and reduced models (G = 15.359, df = 14, p-value ≈ 0.35). Thus, in a statistical sense, there is no advantage of including the additional variables in the model.

The direction of the effects is also the same, which corroborates our hypotheses for these predictors. The effect size (in odds ratios) for the pronominality of the Recipient is approximately the same as in the full model, with a small decrease in the upper limit of the 95% CI. However, the effect of length difference has increased; this result indicates that small length differences tended to have a stronger effect in the reduced model than in the full model. The variance of the

random factor of verb has also increased from 1.695 to 2.392; this finding also indicates that the variation that is associated with specific verbs has increased in comparison with the estimates of the full model.

Recall that the verb is modeled as a random intercept. This means that, all else being equal, certain verbs are more predisposed to take a PP Recipient, whereas other verbs are more likely to take an NP realization. Table 4 provides the estimated intercepts that are associated with each verb in our data sample. A negative sign indicates a preference for a NP realization, and a positive sign indicates a preference for a PP realization.

Table 5: values for the intercepts associated with each verb in the minimal model.

accord	0.205	leave	-0.048
afford	0.002	lend	0.885
allot	0.165	mail	-0.049
allow	-0.957	offer	0.368
assign	0.114	owe	0.902
assure	-1.221	pay	0.009
award	0.013	permit	-1.002
bring	0.656	present	0.270
carry	0.040	promise	-1.053
cause	0.300	read	0.479
charge	-0.158	refuse	0.043
cost	-1.808	sell	1.415

deal	0.073	send	1.212
deliver	0.474	serve	0.304
deny	-0.336	show	-0.800
extend	1.881	slip	0.392
feed	-0.034	submit	1.530
get	-0.911	supply	0.328
give	-1.210	teach	-2.445
grant	-1.057	tell	-2.051
hand	-0.580	wish	-1.009
issue	1.839	write	0.953
lease	0.755		

The next section discusses these values and the other results in comparison with the findings in the literature.

Discussion

Our corpus results indicate that the dative alternation in IndE is largely motivated by three factors: the pronominality of the Recipient, the difference in length between the Recipient and the Theme and verbal semantics. The directions of these effects are largely consistent with our hypotheses based on previous findings. Moreover, the pronominality of the Recipient and length differences were also found to yield the strongest effects in other studies (Wolk, Bresnan,

Rosenbach and Szrmecsányi 2007; Kendall, Bresnan and Van Herk, 2011; de Marneffe, Grimm, Arnon, Kirby and Bresnan, 2012; Theijssen, Bresnan, Ford and Boves subm.).

With regard to the verb, we found that the predicted intercept values ranged from -2.445 (*teach*) to 1.881 (*extend*). The values for the most common verbs (*give*, *tell*, *pay*, *offer*, *show* and *bring*, which together account for 70% of the corpus sample) are provided in Table 6 in comparison with the estimates of Bresnan and Ford (2010).

Table 6: Comparison of the estimated intercepts

Current dataset		Bresnan & Ford (2010)	
give	-1.210	give.a*	-1.3000
		give.c	0.0627
		give.t	-0.1314
tell	-2.051	tell.c	-5.787
pay	0.009	pay.a	5.2559
		pay.t	-1.9805
offer	0.368	offer.a	1.3923
		offer.f	0.7515
show	-0.800	show.c	-0.8993
bring	0.656	bring.a	3.1927
		bring.t	2.0043

* the indexes refer to the semantic class in which the verb was observed: a = abstract; c = communication; t = transfer; f = future transfer.

The directions of the effects (cf. the signs) are nearly identical (*pay* is an exception). In our dataset, *give* was used in either an abstract or a transfer sense (cf. Table 7); therefore, the results are comparable. The same holds true for *tell* (communication), *offer* (abstract and future transfer), and *bring* (abstract and transfer).

Table 7: Bivariate distribution of the six most common verbs according to semantic class

	abstract	communication	transfer	future transfer	Total
give	361	0	72	0	433
tell	0	54	0	0	54
pay	44	0	5	0	49
offer	31	0	0	14	45
show	11	32	0	0	43
bring	22	0	15	0	37

The estimated variance in our dataset was slightly smaller than in the study of Bresnan and Ford (2010): 2.392 vs. 2.5246, respectively. This suggests that the preferences of the verbs is less outspoken in IndE than in American spoken English.

We found no statistical evidence that the other variables under examination are involved. This finding is remarkable considering that these variables have repeatedly been found to play a role in previous studies (if only accounting for less than 10% of the variability in Theijssen, Bresnan, Ford and Boves (subm.).

Previous studies have explained their findings in terms of *harmonic alignment*. The latter term refers "to the tendency for linguistic elements that are more or less prominent on a scale

(such as the animacy or nominal-expression type scales) to be disproportionately distributed in respectively more or less prominent syntactic positions (such as preceding in word order or occupying a superordinate syntactic position)" (Bresnan and Ford 2010: 183). Accordingly, all else being equal, there is a tendency in English to place animate, definite, singular, pronominal and shorter objects (irrespective of Theme or Recipient) before inanimate, indefinite, plural, nominal and longer objects. Our findings are consistent with this tendency, but we appear to have an insufficient number of significant variables to speak of an "alignment".

Our dataset may be insufficient in size to obtain significant results. Alternatively, the dative alternation in IndE may have developed in a manner that differed greatly from the evolution of the other macro-regional varieties of English to the extent that the other variables no longer play a significant role. We would need more data to evaluate this hypothesis. However, the results for pronominality may be explained in terms of discourse status. Pronominal Recipients are generally given in the context of discourse, and it is largely acknowledged that there is a strong tendency in English to place given referents before new information. The effect of length, which is generally acknowledged as a strong motivating factor in constituent ordering, may be explained in terms of processing considerations.

Our results further corroborate earlier findings that the to-dative construction is proportionally more frequent in IndE than in other macro-regional varieties of English. Different explanations have been proposed for this observation.[4] One hypothesis states that IndE preserved the complementation pattern of 19th century BrE (Olavarría de Ersson and Shaw 2003: 158; Muhkerjee and Hoffman 2006: 166~~;~~). However, diachronic evidence pertaining to the dative alternation in BrE based on the ARCHER corpus is inconsistent with this hypothesis. Wolk, Bresnan, Rosenbach and Szrmecsányi (2012: 21) found that "dative proportions are fairly stable

in real time. The share of ditransitive datives modestly fluctuates between 61% (1800-1849) and 70% (1900-1949)." Moreover, comparing the proportion that was found for our sample with that of the 19th century, we found strong evidence that the number of to-datives in our Kolhapur sample was significantly higher than would be expected if there were no difference in proportions between the two samples ($\chi^2(1) = 12.75$, p-value < 0.001).

Table 8: Dative proportions in Kolhapur (late 20th century IndE) and the 19th century component of the ARCHER corpus

	Kolhapur	ARCHER (19th Century)
NP	526 (56%)	556 (53%)
PP	417 (44%)	311 (47%)
Total	943 (100%)	867 (100%)

These findings suggest that the higher proportion of to-dative constructions in IndE is not because IndE retained the proportion of 19th century BrE. Rather, IndE appears to have developed away from BrE.

Why is there a larger proportion of to-datives in IndE than in other varieties of English? Mukherjee and Hoffmann (2006: 154) relate this difference to the tendency in IndE to use *give* as a "compound verb" (e.g., *give explanation, give punishment, give provocation*). Mukherjee and Hoffmann observed that Indian English compound verb constructions typically choose prepositional objects (e.g., *give explanation to the problem*, *give provocation to the deceased*). The use of *give* as a compound verb is in our corpus overly small to draw strong conclusions

regarding this hypothesis. However, we would like to add another argument in favor of Mukherjee and Hoffman's hypothesis.

Compound verbs are one of the features that identify the South Asian "Sprachbund" (Butt 1995; Hook 1974; Steever 1988; Verma 1993). For instance, the Hindi verb *denā* 'to give' is combined with an infinite list of nouns, such as those above, *dard* 'pain', *yād* 'memory', etc. Knowing that most authors of IndE are bilingual and fluent in at least one Indian vernacular language in addition to IndE (Schiffman 2004), constructions that include *to give pain* or *to give punishment* could be "calques" from Hindi; this possibility could explain why such constructions are popular in IndE.

We wish to propose a second explanation (which is also compatible with the first explanation) based on transfer from Hindi. There are notable similarities between the English to-dative construction and the ditransitive construction in Hindi. In Hindi, the Recipient must always be explicitly marked with the postpositional marker *ko*, as illustrated in (17). However, unlike in the English to-dative construction, the Recipient tends to precedes the Theme in Hindi (cf. Koul 2008: 216, the order is only reversed if the Theme is emphasized):

(17) maiṃ [apnī bahan=ko]$_{\text{RECIPIENT}}$ [yah kitāb]$_{\text{THEME}}$ deti hūṃ
I [to my sister]$_{\text{RECIPIENT}}$ [the book]$_{\text{THEME}}$ give

The prepositionless DOC alternative, which is given in (18), is ungrammatical:

(18) *maiṃ [apnī bahan]$_{\text{RECIPIENT}}$ [yah kitāb]$_{\text{THEME}}$ deti hūṃ

I [my sister] [the book]$_{\text{THEME}}$ give

Other Southern Indian languages, including Dravidian Tamil and Telugu, similarly require an explicit dative case marker. Thus, the preference in IndE for a to-dative construction may be related to explicit case marking in Hindi or other Southern Indian languages. Of course, further research is required to test this possible linguistic transfer. For instance, one could compare the preferences of IndE speakers whose first or second language is Hindi with IndE speakers who do not have Hindi as a first or second language. The 100-split task that was performed by Bresnan and Ford (2010) would be an appropriate test.

Conclusion

This study examined the dative alternation in IndE. We were particularly interested in exploring what motivates speakers' use of one of the two orders and hypothesized that the same factors that are known to be associated with the dative alternation in other macro-regional varieties of English would also be involved in IndE. The effects of 14 variables were evaluated by means of a mixed-effects logistic regression analysis on a corpus sample of N = 943 observations that were drawn from the Kolhapur corpus. We found statistical evidence for the effect of three variables, including the pronominality of the Recipient, the difference in length between the Theme and the Recipient, and verb semantics. As expected, the to-dative construction was associated with a pronominally expressed Recipient and with a Recipient that is longer than the Theme. We also found that certain (e.g., *bring* and *offer*) verbs are more inclined to be used with the to-dative

construction than other verbs. The direction and strengths of the significant effects were similar to those found in previous studies on the dative alternation. Remarkably, however, we did not find evidence for the effects of the other variables under analysis. The latter problematizes the explanatory role of the concept of *harmonic alignment*. We argued that our results are best explained in terms of discourse status (pronominal objects are highly correlated with given referents) and processing constraints (longer constituents generally occur later in clauses). Our study also corroborates earlier findings that the to-dative construction is more frequent in IndE. We argued that this finding indicates that IndE developed away from BrE. We proposed that this tendency may be influenced by the use of "compound verbs" and prepositional ditransitive patterns in Hindi and Southern Indian languages, although more research is required to support the latter hypotheses. We like to end with some suggestions for future research. Possible topics of interest include the role of modality (i.e., spoken vs. written) and formality.

[1] There are some notable exceptions to the main features of the dative alternation. First, the sentence verb may denote the denial of a transfer or the lack thereof (e.g., with the verb *deny* or *refuse*). Second, the object order of both constructions may be reversed in specific contexts ("heavy noun phrase shift") (cf. a) or because of dialectal preferences (Gast 2007) (cf. b); both examples are taken from Huddleston and Pullum (2002: 248, 250). In this study, however, we focus on the canonical orders.

a) He gave to charity everything he earned from the concert.

b) He gave it her.

[2] More information on the Kolhapur corpus can be found at: http://corp.hum.ou.dk/itwebsite/corpora/corpman/KOLHAPUR/INDEX.HTM <DOA 10 Jun. 12>. The fact that the Kolhapur corpus consists of written texts makes it more suitable to compare with 19th century British English than the more recent ICE-Indian corpus, which also includes spoken texts.

[3] Several two-way interactions were evaluated unsuccessfully. We also fitted verb sense as a random factor, but the estimated variance was lower than with the verb, which is why we chose to retain the latter instead. All statistical analysis were performed in R.

[4] A more "speculative" explanation for the frequency of the to-dative construction in IndE has been suggested by Olavarria de Ersson and Shaw (2003: 159), who argue that this frequency may reflect a cultural difference. In their view, South Asian cultures would have a more holistic view of the world, in which an individual is regarded as being part of the larger whole rather than at the center of the world, as would be the case in Northern European cultures. This difference would be reflected in the IndE preference to use the to-dative construction, which would profile the Theme rather than the Recipient. Much more evidence would be required (particularly for the alleged cultural differences) before one could seriously consider this explanation.

References

Arnold, Jennifer E., Losongco, Anthony, Wasow, Thomas, & Ginstrom, Ryan (2000) Heaviness vs. Newness: The Effects of Structural Complexity and Discourse Status on Constituent Ordering. *Language, 76*(1), 28–55.

Baayen, R. Harald (2008) *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge, UK ; New York: Cambridge University Press.

Bock, J. Kathryn and Irwin, David E. (1980) Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 467–484.

Bock, J. Kathryn, Loebell, Helga and Morey, Randal (1992) From conceptual roles to structural relations: bridging the syntactic cleft. *Psychological Review, 99*, 150–171.

Bresnan, Joan W. and Ford, Marilyn (2010) Predicting Syntax: Processing Dative Constructions in American and Australian Varieties of English. *Language, 86*(1), 168–213.

Bresnan, Joan W., Cueni, Anna, Nikitina, Tania and Baayen, R. Harald (2007) Predicting the Dative Alternation. In Gerlof Bourne, Irene Kraemer & Joost Zwarts (Eds.), *Cognitive Foundations of Interpretation* (pp. 69–94). Amsterdam: Royal Netherlands Academy of Science.

Bresnan, Joan W. and Ford, Marilyn (2010) Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language, 86*(1), 168–213.

Bresnan, Joan W. and Nikitina, Tania. (2007) The Gradience of the Dative Alternation. In Linda Uyechi and Lia Hee Wee (Eds.), *Reality Exploration and Discovery: Pattern Interaction in Language and Life*. Stanford: CSLI Publications.

Butt, Miriam (1995) *The structure of complex predicates in Urdu*. Stanford: CSLI Publications.

Collins, Peter (1995) The indirect object construction in English: an informal approach. *Linguistics, 33*(1), 35–49.

de Marneffe, Marie–Catherine, Grimm, Scott, Arnon, Inbal, Kirby, Susannah and Bresnan, Joan W. (2012) A statistical model of the grammatical choices in child production of dative sentences. *Language and Cognitive Processes, 27*(1), 25–61.

Erteschik–Shir, Nomi (1979) Discourse Constraints on Dative Movement. In Talmy Givón (Ed.), *Syntax and Semantics.* (pp. 441–467). New York: Academic Press.

Garretson, Gregory (2003) *Optimal typology of Determiner Phrases coding manual*. Boston: Boston University.

Gast, Volker (2007) I gave it him — on the motivation of the 'alternative double object construction' in varieties of British English. *Functions of Language, 14*(1), 31–56.

Givón, Talmy (1984) *Syntax. A Functional-Typological Introduction. Vol. I.* Amsterdam: Benjamins.

Green, Georgia M. (1974) *Semantics and syntactic regularity*. Bloomington (Ind.): Indiana university press.

Gries, Stephan T. (2005) Syntactic priming: A corpusbased approach. *Journal of Psycholinguistic Research, 34,* 365–399.

Halliday, Michael A. K. (1970) Language structure and language function. In John Lyons (Ed.), *New horizons in linguistics* (pp. 140–165). Harmondsworth, Middlesex: Penguin.

Hawkins, John A. (1994) *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Hook, Peter E. (1974) *The compound verb in Hindi*. Michigan: Center for South and Southeast Asian Studies.

Huddleston, Rodney D. and Pullum, Geoffrey K. (2002) *The Cambridge grammar of the English language*. Cambridge, UK ; New York: Cambridge University Press.

Kendall, Tyler, Bresnan, Joan W. and Gerard Van Herk (2011) The Dative Alternation in African American English: Researching Syntactic Variation and Change across Sociolinguistic Datasets. *Corpus Linguistics and Linguistic Theory*, *7*(2), 229–244.

Koul, Omkar N. (2008) *Modern Hindi Grammar*. Springfield: Dunwoody Press.

Lapata, Maria (1999) Acquiring lexical generalizations from corpora: A case study for diathesis alternations. *Proceedings of the North American Chapter of the Association for Computational Linguisics. 37* (pp. 397–404).

Levin, Beth (1993) *English verb classes and alternations: a preliminary investigation*. Chicago (Ill.): University of Chicago press.

Muhkerjee, Joybrato and Hoffman, Sebastian (2006) Describing verbcomplementation profiles of New Englishes. A pilot study of Indian English. *English WorldWide, 27*(2), 147–173.

Nihalani, Paroo, Tongue, Ray K. and Hosali, Priya (1979) *Indian and British English. A Handbook of Usage and Pronunciation*. Delhi: Oxford University Press.

Oehrle, Richard T. (1976) *The grammatical status of the English dative alternation.* Unpublished PhD thesis, Cambridge (MA), MIT.

Olavarría de Ersson, Eugenia and Shaw, Philip (2003) Verb complementation patterns in Indian Standard English. *English World–Wide, 24*(2), 137–161.

Ozón, Gabriel (2009) *Alternating ditransitives in English: a corpus–based study*. Doctoral thesis, UCL (University College London).

Pinker, Steven (1989) *Learnability and cognition: the acquisition of argument structure*. Cambridge (MA): MIT Press.

Prat–Sala, Mercè and Branigan, Holly P. (2000) Discourse constraints on syntactic processing in language production: A crosslinguistic study in English and Spanish. *Journal of Memory and Language, 42*(2), 168–182.

Ransom, Evelyn (1979) Definiteness and animacy constraints on passive and double–object constructions in English. *Glossa, 13*, 215–240.

R Development Core Team (2011) R: A Language and Environment for Statistical Computing. Vienna. Austria. http://www.Rproject.org

Schiffman, Harold (2004) *Bilingualism in South Asia: Friend or Foe?* Paper presented at the Proceedings of the International Symposium on Bilingualism (ISBL4), Tempe: AZ.

Smyth, Ronald H., Hogan, John T. and Prideaux, Gary D. (1979) The Effect of Context on Dative Postition. *Lingua, 47*, 27–42.

Snyder, Kieran (2003) *The relationship between form and function in ditransitive constructions.* University of Pennsylvania, Philadelphia.

Steever, Sanford (1988) *The serial verb formation in the Dravidian languages*. Delhi: Motilal Banarsidass.

Theijssen, Daphne (2008) Using the ICEGB Corpus to model the English dative alternation. *Online Proceedings of the Aston Postgraduate Conference on Corpus Linguistics*.

Theijssen, Daphne (2009) Variable selection in Logistic Regression: The British English dative alternation. *Online Proceedings of the 14th Student Session at the 21st European Summer School in Logic, Language and Information (ESSLLI09)*, 85–95.

Theijssen, Daphne, Bresnan, Joan, Ford, Marilyn and Boves, Lou (subm.). In a land far far away... A probabilistic account of the dative alternation in British, American and Australian English.

Thompson, Sandra (1995) The iconicity of "dative shift" in English: Considerations from information flow in discourse. In Marge E. Landsberg (Ed.), *Syntactic Iconicity and Linguistic Freezes. The Human Dimension* (pp. 155–175). Berlin/New York: Mouton de Gruyter.

Thompson, Sandra and Koide, Yuka (1987) Iconicity and indirect objects in English. *Journal of Pragmatics, 11*(3), 399–406.

Verma, Manindra (1993) *Complex predicates in South Asian languages*. Delhi: Manohar.

Wasow, Thomas (2002) *Postverbal behavior*. Stanford (CA): CSLI Publications.

Wolk, Christopher, Bresnan, Joan, Rosenbach, Anette and Szrmecsányi, Benedikt (2012) Dative and genitive variability in Late Modern English: Exploring crossconstructional variation and change. *Diachronica* (forthcoming).