

Chapter 1

Givenness marking in a mixed system: Constituent order vs. determiners

Alexandra Simonenko

Research Foundation Flanders & Ghent University

Anne Carlier

Sorbonne University, EA STIH

This paper investigates the interaction between constituent order and the use of determiners as means of marking givenness, understood here as a non-presuppositional existential inference that arises as a result of interpreting a predicate with respect to a context-specified situation, in light of (a version of) *New \succ Given principle of Kučerová (2012). We attribute the principle to how situation binding operates in clauses, instead of postulating a presupposition-introducing operator and test it on new quantitative data from Medieval French, a system employing both determiners and constituent order for information structuring. Our results show that the constraint in question is respected across the board except for the cases when it is obviated by the presence of a morphological trigger of existential presupposition. We also show that a game-theoretic simulation incorporating this constraint matches very closely historical French data.

1 Introduction

This chapter focuses on the interactions between determiner types and constituent order in the marking of givenness in the history of French, on the basis of data from the twelfth to the seventeenth century. We understand givenness here in a weak sense of an existential inference that emerges when a nominal predicate is interpreted relative to a particular, discourse-specified situation. Medieval French is commonly assumed to have employed syntactic means for the expression of information structure. Because of the absence of native speaker judgments and speech recordings for historical data, identifying the exact informa-



tion structural import of syntactic configurations no longer available in Modern French, such as the clause with a preverbal object in (1), is a complex task. However, the consensus states that the variable placement of arguments had mostly to do with the way their denotation related to the contextual information.

- (1) Iceles miracles vit li pelerins
these miracles saw the pilgrim
“The pilgrim saw these miracles”

(1210-BORON-PENN-P,32,441)

Marchello-Nizia (1995), who was among the first to investigate the tendency of Old French to have the verb immediately follow the first constituent (i.e. the so called V2-order, explored in a long series of works starting with Skårup (1975)), suggested that the first preverbal position was reserved for elements establishing a link with the previous discourse (for a similar intuition in Vennemann (1974) and Harris (1978)). Rinke & Meisel (2009: 117) argue that “the pre-verbal [subject] position correlates with a topic-interpretation and the post-verbal position with a non-topic interpretation”. Kaiser & Zimmermann (2011: 24) propose that “the positioning of one or more non-subject constituents to the left of the finite verb in declarative root clauses directly correlates with their discourse status, i.e. with their interpretation as either topicalized or focalized constituents”. They assume a split CP involving Topic and Focus projections. Based on an extensive corpus data analysis, Labelle & Hirschbühler (2018) conclude that although the distribution is not categorical, the initial constituent in V2 configurations tend to be topical. V2 with non-subject preverbal constituents progressively becomes more rare until the constituent order in declarative sentences effectively converges onto SVO.

At the same time, French already has *le/la/les* determiners, the frequency of which will increase over the course of history. These determiners have to be analysed as definiteness markers. The two series of phenomena, constituent order and determiners are closely related since both are crucially involved in structuring the propositional content with respect to the background information. The idea that flexible word order and determiners in stand in some sort of a “trade-off” relation with each other has emerged in the literature on historical Romance languages on multiple occasions (e.g. Vincent:1997, Ledgeway:2012).

In this paper, we do not commit ourselves to any claims about causes of word order changes in French. Rather, we assume that word order changes happen for independent prosodic and/or morphological reasons (such as the loss of case marking). We also, strictly speaking, do not present evidence that there is a

causal relation between word order changes and changes in the use of determiners. Rather, we provide supporting arguments for such claims.

Definite, possessive, and demonstrative determiners are assumed in the Frege/Strawson tradition to function as existential presupposition triggers since their felicitous use requires the background to entail the existence of an individual or entity with certain properties. The English utterance in (2) where the subject DP is headed by a definite determiner is felicitous just in case the existence of a dog in some relevant domain is part of the participants shared knowledge.¹

- (2) The dog is barking.

The increase in the frequency of definite determiners is closely followed by the increase in frequency of indefinite determiners, as discussed in Carlier:2013, which signal that a novel referent is being introduced and that a definite determiner could not have been used (Heim 1982, Heim 1991).²

Conditions on the use of definite and indefinite determiners in English in some other languages partially correspond to the conditions on argument ordering. For instance, this is the case in Czech.³ Consider examples 3–4 from Kučerová (2012), where a clause-initial argument is likely to be interpreted as denoting an entity whose existence is part of the background information, whereas a postverbal argument is likely to be interpreted as introducing a novel referent.

- (3) Chlapec našel lízátko.
boy.Nom found lollipop.Acc
“The/a boy found a lollipop.” #“A boy found this lollipop.”
- (4) Lízátko našel chlapec.
lollipop.Acc found boy.Nom
“A boy found this lollipop.”

Kučerová (2012) argues that in Czech object scrambling is a means of aligning the syntactic structure with the (default) Given \succ New order, where a constituent is considered as given if it has an antecedent in the preceding discourse and if the

¹ Even in light of the analyses that deny the definite determiner an existential presupposition, such as Coppock & Beaver (2015), as will be discussed below, it is enough for our purposes that in most cases they give rise to an existential inference as a result of the nominal predicate being interpreted with respect to a contextually provided situation.

² SimonenkoCarlier:review give quantitative data on the changes in the determiner system in French over the course of time.

³ Titov (2012) makes a similar point for Russian.

context entails the existence of an entity with the property denoted by this constituent. On this proposal, the two grammatical phenomena related to existential presupposition marking – syntactic and morphological (i.e. via pronominalization or a determiner) – can substitute for each other. The evolutionary trajectory of diachronic French data makes it an ideal test-case for this hypothesis since for several centuries the available texts feature both a very flexible and evidently information structure-driven constituent order *and* emerging definite determiners.

We will claim that medieval French data corroborate (an amended version of) Kučerová's (2012) model which predicts the infelicity of *New \succ Given order within a propositional domain. We will show that all bare noun configurations involve Given \succ New sequence, and that determiners have an obviating effect on this principle in that New \succ Given is possible if the second argument involves a presupposition-triggering determiner.

Based on this proposal, we build a game-theoretic simulation of the interpretation of a class of utterances and show that the results of the simulation are almost identical to the empirically observed picture in historical French.

We also show that the *New \succ Given makes a correct prediction with regard to the relative frequencies of different constituent orders. Finally, we use determiner distribution patterns to identify the information structural import of the OVS order, the only non-marginal configuration involving O \succ S. We show that objects in OVS occur exceptionally frequently with demonstratives, which we analyse as signalling their status as shifted topics. Our quantitative analysis is based on the treebanks MCVF and Penn Supplement to MCVF.⁴

In the next section we discuss Kučerová's model, propose an amendment and lay out the predictions the amended model makes for the historical French data. In section 3 we show these predictions to be borne out. Section 4 presents our Rational Speech Act model-based simulations and compares its predictions to the historical French data. In section 5 we discuss the explanatory potential of our model of givenness marking for the constituent order frequencies. Section 6 concludes.

2 Marking givenness

Using Modern Czech as an empirical base, Kučerová (2012) formulates the information flow principle in (5), which states that a constituent interpreted as con-

⁴ MCVF and Penn Supplement to MCVF with about 1,5 million words is the largest treebank for French diachrony to date.

veying new information cannot precede a constituent interpreted as conveying given information.

(5) Generalization *New \succ Given

Within a domain $[_{Dom} Y \dots X]$, if X is given, so is Y. Kučerová (2012: 14)

The generalization captures the range of possible interpretations for utterances in terms of the sequences of new and given information. For instance, for (3), it captures the unavailability of the interpretation whereby *chlapec* (“boy”) is interpreted as new and *lízátko* (“lollipop”) as given. It also captures the fact that if *lízátko* (“lollipop”) is made to precede *chlapec* (“boy”), the given interpretation becomes available for the former, as (4) shows.

Kučerová (2012: 18) assumes the notion of givenness as spelled out by Schwarzschild (1999: 151), as in (6).⁵

(6) An utterance U counts as Given iff it has a salient antecedent A and

- a. if U’s type is e, then A and U corefer;
- b. otherwise, modulo \exists -type shifting, A entails the Existential F-closure of U.⁶

Kučerová (2012: 14) derives this generalisation from the mechanism of givenness marking in natural language. Specifically, she proposes that a givenness presupposition can be triggered by a syncategorematic G(ivenness) operator that can be applied anywhere in a propositional domain, dividing the domain into given (higher) and new (lower) parts, as illustrated in figure 1 from Kučerová (2012: 3).⁷

⁵ Kučerová (2012: 18) notes that she follows Sauerland 2005 in assuming that “givenness gives rise to an existential presupposition”, without spelling out the details”.

⁶ Existential F-closure involves replacing focused expression by existentially closed variables.

⁷ For the technical details of the recursive application of the G-operator which introduces restrictions on arguments’ domains we refer the reader to Kučerová (2012) and (in even greater detail) Šimík & Wierzbica (2015).

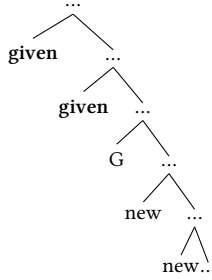


Figure 1: Partitioning of a clause by the Givenness operator

Kučerová (2012) also assumes that the insertion of such an operator is necessary if the presupposition is satisfied in a given context and if there are no morphological means of marking it in the numeration (in the Minimalist sense). The latter part is based on the Maximize Presupposition! principle of Heim (1991) supplemented with an assumption that the competition takes places between structures generated from the same numeration.

The predicted infelicity of New \succ Given sequences corresponds to a presupposition failure since the operator G has the effect that all constituents to its left carry a givenness presupposition. One consequence of this proposal is that configurations where new material linearly precedes old in a given domain are only felicitous if the given material is morphologically marked as such. Kučerová (2012) assumes that once the givenness presupposition is morphologically marked, the G operator is not inserted. Morphological triggers of givenness presupposition involve proper nouns and personal pronouns.

Based on experimental results for Czech, Šimík & Wierzbą (2015) replace the *New \succ Given principle with a *Non-presupposed \succ Presupposed constraint. They note, however, that the constraint is not absolute in that its violation does not result in the same degree of infelicity as the use of *a* instead of *the* in English in a context suitable for the latter.

We build on this version, proposing that the relevant notion is existential non-presuppositional inference rather than a hard presupposition and that its violation causes a downstep in acceptability rather than strong infelicity. We also add to it an obviation condition that morphological triggers of existential presupposition, such as personal pronouns, proper names, demonstrative, definite, and possessive determiners, are exempt from the constraint. In addition to arguments that involve morphological markers of existential presupposition, we note that an argument can be exempt from existential opposition altogether, as in the case of bare nouns forming complex predicates with finite verbs. We as-

sume that this applies to idiomatic expressions or light verb constructions such as *avoir nom* (“to be called”), *avoir peur* (“to be afraid”), *avoir cure* (“to need”), *faire mal* (“to hurt”), where the bare noun cannot be interpreted as a referential expression. An example of such construction is given in (7).

- (7) out num_{obj} cil_{sbj} Nabal
had name this Nabal
“this (man) was called Nabal” (1150-QUATRELIVRE-PENN-P,49.1841)

An amended version of the constraint is given in (8) where the relevant notion of givenness is defined as in (9).

- (8) Generalisation #New \succ Given
Within a domain [_{Dom} Y ... X], if X is given, so is Y
unless X involves a morphological trigger of existential presupposition
or unless the non-presupposed/presupposed opposition does not apply to
one of the arguments,
- (9) A constituent *C* of an utterance *u* in a context *c* (in Stalnaker’s sense)
interpreted with respect to a situation *s* is considered given if *c* entails the
non-emptiness of the extension of *C* in *s*.

Instead of Kučerova’s syncategorematic introducer of domain restrictions (i.e. presupposition trigger) *G*, we assume that the relevant operator is a situation binder \sum_G , which binds situation variables of predicates to a topic situation down to a point where there is another binder. That is, \sum_G binds all unbound situation variables. Following Kratzer (2017) and Schwarz (2009: 127–133), we assume that topic situations can be derived from Questions Under Discussion (Roberts (1996), Büring (2003)). Specifically, a topic situation is a minimal situation that exemplifies the set of situations in which the answers to QUD are the same as in the actual world.⁸

The relevant existential inference, which constitutes the content of givenness on our account, is an inference that the extensions of predicates interpreted relative to a topic situation are non-empty. The inference arises because the topic situation is a situation whose contents are likely known to the speech act participants. That the relevant existential inference in question is non-presuppositional

⁸ Definitions of exemplification and minimality from Kratzer (2017): A situation *s* exemplifies a proposition *p* iff whenever there is a part of *s* in which *p* is not true, then *s* is a minimal situation in which *p* is true. A situation is a minimal situation in which a proposition *p* is true iff it has no proper parts in which *p* is true.

on our account matches the conclusion reached in Šimík & Wierzba (2015) that their *Non-presupposed \succ Presupposed constraint is relatively mild. If we assume that interpreting a predicate relative to a topic situation gives rise to an existential inference because the speech act participants expect the contents of the topic situation to be known, the inference is cancellable to the extent that this expectation can prove wrong, that is, that in some cases a topic situation does involve entities whose existence is not part of the common ground.

In the absence of any other sources of values for situation variables, this derives, in particular, the $\#NP1_{New} \succ NP2_{Given}$ constraint, since if the situation argument of NP2 is bound, that of NP1 is bound as well (and is therefore interpreted relative to the topic situation, giving rise to an existential inference), just because the binding operators in an across-the-board fashion from top to bottom.

With regard to the rationale behind the obviation conditions, the presence of a determiner introducing its own resource situation pronoun and triggering existential presupposition relative to that situation (as, we assume, possessive, demonstrative, and definite determiners do) makes \sum_G qua the binding mechanism unnecessary (although the determiner's situation pronoun may co-refer with the topic situation, cf. Schwarz (2009)). As a result, a noun phrase interpreted with respect to some other situation may precede a noun phrase with a presupposition-triggering determiner. This is illustrated in figure (2) where s_{topic} is a topic situation pronoun, s_r is a resource situation pronoun associated with a determiner, -s stands for a situation argument of a nominal predicate, and \sum_G is the situation binder in question. In this configuration, the two upper NPs are interpreted with respect to the topic situation, whereas the situation argument of the lowest NP is valued by a separate resource situation introduced by a determiner.

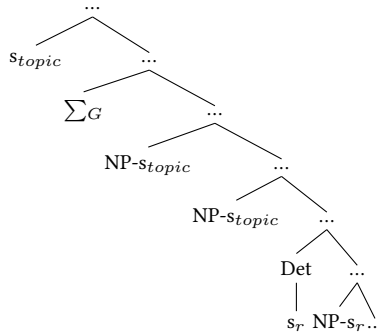


Figure 2: Givenness operator as a situation binder

2.1 Morphological triggers of existential presupposition

We assume the Logical Forms of demonstrative, definite, and possessive determiners involve a resource situation pronoun which “stops” the binding triggered by \sum_G . The LFs and lexical entries for definite and demonstrative determiners are based on Heim (2011), Elbourne (2008), and Schwarz (2009). The entry for possessives is based on Simonenko & Carlier (2019). All these are given in (10)–(15) for the sake of concreteness.

- (10) LF of a definite determiner: $[[D_{def} s_r] NP]$
- (11) $[[D_{def}]] = \lambda s_\sigma. \lambda P_{\langle e, \sigma t \rangle} : \exists x \forall y [Max(P)(y)(s) \rightarrow x = y] . \iota x [Max(P)(x)(s)],$
where $Max(P) = \lambda x_e. \lambda s_\sigma . P(x)(s) \ \& \ \neg \exists y [P(y)(s) \ \& \ x < y]$
- (12) LF of a demonstrative determiner: $[i [[D_{dem} s_r] NP]],$
where i is the index of a silent individual pronoun
- (13) $[[D_{dem}]] = \lambda s_\sigma. \lambda P_{\langle e, \sigma t \rangle} . \lambda y_e : \exists x [P(x)(s) \ \& \ x = y] . \iota x [P(x)(s) \ \& \ x = y]$
- (14) LF of a possessive determiner: $[i_{poss} [[D_{poss} s_r] NP]],$
where i is the index of a silent individual pronoun
- (15) $[[D_{poss}]]^{c,g} = \lambda s_\sigma. \lambda P_{\langle e, \sigma t \rangle} . \lambda y :$
 $\exists! x [\lambda s_\sigma . \lambda y_e . \lambda z_e . z \text{ belongs to } y \text{ in } c \ \& \ P(x) \text{ in } s)(s)(y)(x)] .$
 $\iota x . Max(\lambda s_\sigma . \lambda y_e . \lambda z_e . z \text{ belongs to } y \text{ in } c \ \& \ P(y) \text{ in } s)(s)(y)(x))$

Because of the existential presupposition they involve, these entries, if used in a felicitous sentence, give rise to an existential inference. The resource pronoun in their logical forms, in the absence of external binders, does not propagate its value beyond the local DP, which predicts the felicity of $NP1_{New} \succ NP2_{Given}$ if the existential inference results from the use of a determiner. In particular, it correctly predicts the felicity of the Czech example in (16), from Kučerová (2012).

- (16) Chlapec našel **ten** lízátko.
boy.Nom found this lollipop.Acc
“A boy found this lollipop.”

2.2 Predictions

With respect to its constituent order flexibility, Old French is more similar to Modern Slavic languages than to Modern French. For instance, a transitive clause

with a nominal subject and object can have any of the 6 possible constituent orders: SOV, SVO, VSO, OSV, OVS, VOS. Relative frequencies of different orders are given in table 3.

The counts in the table, which we extracted from MCVF (2010) and Penn Supplement to MCVF (Kroch & Santorini 2010), include all finite transitive clauses, both matrix and embedded, with nominal subjects and objects.⁹ We excluded all cases of pronominalization, first, because of their often restricted syntactic distribution in comparison with nominal arguments and, second, because pronouns either trigger existential presupposition or are explicitly incompatible with it and therefore will not help us evaluate the #New \succ Given principle.

Table 1: Constituent order in transitive clauses

	OSV	OVS	SOV	SVO	VOS	VSO
XI c.	0.02 (2)	0.13 (17)	0.14 (18)	0.62 (83)	0.02 (3)	0.05 (6)
XII c.	0.01 (27)	0.11 (203)	0.12 (219)	0.61 (1120)	0.05 (95)	0.09 (173)
XIII c.	0.00 (3)	0.04 (23)	0.02 (13)	0.77 (493)	0.02 (15)	0.15 (97)
XIV c.	0.00 (3)	0.03 (37)	0.03 (37)	0.73 (1043)	0.03 (47)	0.18 (255)
XV c.	0.00 (0)	0.02 (11)	0.01 (8)	0.88 (615)	0.02 (13)	0.07 (52)
XVI c.	0.00 (0)	0.02 (5)	0.00 (0)	0.91 (286)	0.02 (6)	0.06 (18)

We are specifically interested in the orderings between subjects and objects. Ignoring verbal position, we give relevant counts and relative frequencies in table 2.

Table 2: Nominal argument order in transitive clauses

Period	OS	SO
XI c.	0.17 (22)	0.83 (107)
XII c.	0.18 (325)	0.82 (1512)
XIII c.	0.06 (41)	0.94 (603)
XIV c.	0.06 (87)	0.94 (1335)
XV c.	0.03 (24)	0.97 (675)
XVI c.	0.03 (11)	0.97 (304)

The conditional version of the #New \succ Given principle in (8) makes a number

⁹ These treebanks, which add up to about 1,5 million words, are described in detail, in particular, in Martineau (2008) and in Simonenko et al. (2018).

of non-trivial predictions. Specifically, given a transitive clause with overt nominal subject and object, we expect to find the orders in (17) but not in (18), where DET stands for a morphological trigger of existential presupposition.

(17) Predicted licit patterns for sequences involving new and old material:

- | | |
|--|-------------------------|
| A1 (DET-)S _{given} O _{new} | |
| A2 (DET-)O _{given} S _{new} | |
| A3 S _{new} DET-O _{given} | (obviated #New > Given) |
| A4 O _{new} DET-S _{given} | (obviated #New > Given) |

(18) Predicted illicit patterns for sequences involving new and old material:

- B1 S_{new} O_{given}
B2 O_{new} S_{given}

3 Testing the predictions

In the previous section we outlined the predictions made by *New > Given supplemented with a proviso about the obviating effect of morphological presupposition triggers or arguments to which new/given distinction does not apply, such as incorporated nominals. These predictions are testable in a corpus to the extent that it is representative and that we can approximate infelicity/ungrammaticality of a pattern by the absence thereof in a sufficiently large dataset. Assuming the #New > Given principle, in historical French we expect not to find any patterns where an argument denoting new information linearly precedes an argument denoting given information, unless the second argument features a morphological presupposition trigger or one of the arguments is exempt from new/given opposition. Below we report the results of querying for the licit and illicit configurations listed above.¹⁰

3.1 SO

We first check all the SO clauses which can potentially contain an illicit string B1, that is, finite clauses where both subject and object are bare nouns. There are 283 such clauses in the corpus. Examples (19)–(23) illustrate SO strings involving arguments without morphological triggers of existential presupposition.

¹⁰ We considered transitive clauses with nominal arguments that are not preceded by any of the following: definite, demonstrative, indefinite, possessive or partitive determiner.

- (19) que piété_S venqui paor_O.
that piety conquered fear
“that piety conquered fear.”¹¹ A1
- (20) Juvente bien endoctrinee_S Aporte viellesce senee_O;
youth well instructed brings old.age wise
“Education in the young age brings wisdom in the old age;”¹² A1
- (21) Si con malades_S desirre santé_O
so as sick desires health
“so as a sick man desires health”¹³ A1
- (22) fiebles hum_S dreit_O mais ne conquestast
weak man justice never NEG would.win
“a weak man would never obtain justice”¹⁴ A1
- (23) CASTOR_S en ceste vie Saint ume_O signifie
beaver in this life saint man signifies
“The beaver signifies a saintly person in this life”¹⁵ A1

In the corpus we found no SO strings violating *New \succ Given (i.e. pattern B1 in (18)).

On the basis of (19)–(23), one could argue that the absence of New \succ Given sequences is an epiphenomenon of the sample involving bare nouns only. Namely, it is theoretically conceivable that in Medieval French all noun phrases without determiners either denote abstract notions (as in (19), (20)), receive kind interpretation (as in (23)) or are used in generic statements (as in (20) and (21)). We assume that these interpretations inherently involve an existential presupposition.¹⁶ This means that SO strings involving only those cannot in principle feature New \succ Given sequence and therefore such a sample cannot be used to evaluate the relevant predictions. However, although there are indeed many bare

¹¹ (1155-ENEAS1-BFM-R,87.1898)

¹² (1183-ADGAR-BFM-R,265.3473)

¹³ (1185-COBE-BFM-R,3.28)

¹⁴ (1173-BECKET-BFM-R,74.2014)

¹⁵ (1128-BESTIAIRE-BFM-R,43.564)

¹⁶ We assume that generic statements involve an implicit quantifier over situations coupled with a presupposition of the non-emptiness of its domain, that is, that there exist situations accessible from the evaluation situation in which there are individuals with the nominal property (Lee (1995), von Stechow (1996)).

arguments involving abstract, kind denoting, and generic NPs, there are also cases of SO with bare arguments denoting individuals, as (24)–(26) illustrate.

In (24) the subject *Osbercs e helmes* ('hauberks and helmets') is part of the given information, that is, the context entails that there exists individual with the relevant properties in a prominent situation.

- (24) *Osbercs e helmes_S i getent grant flabur_O*
hauberks and helmets there throw.off great flames
"hauberks and helmets throw off great flames"¹⁷ A1

In (25) the subject is a specific indefinite (the narrator is talking here about Saint Mary who restores humanity to life through Jesus Christ and about Eve, who brings death through sin), and the object is arguably indefinite as well (to be understood as 'a new life').

- (25) *Fame_S vie_O nous restora*
woman life to.us restores
"a woman restored us to life"¹⁸ A1

Example 26 speaks about Christians in some prominent situation.

- (26) *De vus unt crestiens_S cumfort_O.*
from you have Christians comfort
"Christians receive comfort from you."¹⁹ A1

Examples in (27)–(30) are cases of obviation where the presence of a morphological presupposition trigger arguably obviates the #New > Given principle, as predicted in (17).

- (27) *é rasur_S ne li munterad le chief_O*
and razor not him mounted DET head
"and a razor did not touch his head"²⁰ A3
- (28) *grans multitudene d' angeles_S recurent l' anrme_O*
great multitude of angels received DET soul
"a great multitude of angels received the soul"²¹ A3

¹⁷ (1100-ROLAND-V,137.1820)

¹⁸ (1190-BORON-PENN-R,27.431)

¹⁹ (1183-ADGAR-BFM-R,181.1981)

²⁰ (1150-QUATRELIVRE-PENN-P,5.47)

²¹ (ID 1200-SERMMADN-BFM-P,18.136)

- (29) .VII. milie graisles_S i sunent la menee_O
seven tusand bugles there sound DET charge
“seven thousand bugles sound the charge”²² A3
- (30) Si passerent toutes gens d’armes et aultres_S la grose riviere de
so passed all people of army and others DET big river of
la Geronde_O
DET Geronde
“so all the army and others passed the great river of Geronde”²³ A3

3.2 OS

There are 37 finite clauses with OS order where both subject and object are nouns without determiners (but possibly with quantifiers or modifiers) in the corpus we used. A configuration involving Object \succ Subject violates *New \succ Given if the Object is new and Subject is given. The example in (35) may look like a potential violation of #New \succ Given as the object *duel* ‘sorrow’ precedes a clearly given subject *pere et mere* ‘father and mother’.

- (31) n’ert mervolle se duel_O menoient pere et mere_S
not was miracle if mourning led father and mother
“it was not surprising if father and mother were mourning”²⁴ A2

However, *duel* is mentioned just a couple clauses before and, hence, cannot be considered as evoking a new referent. The relevant clause is given in (36).

- (32) duel_O ot li rois_S quant il la voit
mourning had DET king when he her saw
“the king started mourning when he saw it”²⁵ A2

Like *avoir faim* (“to be hungry”, literally, “to have hunger”), *avoir deuil* (‘to suffer’) is a light verb construction combining a copula and a noun that denotes an event or a state. We assume that for these cases the notion of givenness or existential presupposition is undefined and, consequently, the principle does not apply.

²² (ID 1100-ROLAND-V,112.1445)

²³ (1376-FROISSART-7-P,763.1956)

²⁴ (1155-ENEAS2-BFM-R,12.199)

²⁵ (ID 1155-ENEAS2-BFM-R,12.197))

Another example of OS with bare arguments is given in (37), where both NPs are given.

- (33) Force de deïté_O Demustre piz quaré_S
force of divinity shows forequarters
“The forequarters (of a lion) symbolize the divine power.”²⁶ A2

We find the same configuration in (34), where the denotations of both the subject and the object belong to the given information since the relevant passage describes an army being set in motion.

- (34) E destendent acubes_O serjant e escuier_S.
and take.down tentes sergeants and esquires
“And the sergeants and esquires take down the tents.”²⁷

The #New > Given principle predicts an obviation for OS strings where a new object precedes a given subject with a presuppositional determiner (pattern A4 in (17)). This case is illustrated by in (35)–(37).

- (35) Mult grant venjance_O en prendrat l’ emperere_S.
very big revenge of.it will.take DET emperor
“The emperor will take a great revenge of it.”²⁸ A4

- (36) Granz curs_O unt fait li pelerin_S,
big journey have done DET pilgrims
“The pilgrims have done a big journey,”²⁹ A4

- (37) Onbre_O li fet li plus biax arbres_S c’ onques poïst former
shadow him did DET most beautiful tree that ever could form
Nature.
nature
“That most beautiful tree that that the Nature could form gave him
shadow.”³⁰ A4

²⁶ (1128-BESTIAIRE-BFM-R,3.31)

²⁷ (1175-FANTOSME-BFM-R,48.514)

²⁸ (1100-ROLAND-V,112.1449)

²⁹ (1120-BRENDAN-R,59.776)

³⁰ (1170-YVAIN-R,12.383)

As an interim conclusion, in a transitive clauses with bare arguments we found no cases violating #New \succ Given (with obviation conditions), that is, involving an argument associated with given information following an argument not associated with given information. These results are to be only taken as suggestive since the absence of a pattern in a limited sample cannot be straightforwardly interpreted as signalling ungrammaticality. However, the fact that among 320 clauses with bare arguments there are no instances violating #New \succ Given is likely not a matter of chance. That is, we found no patterns B1 and B2. To test this, we compared the number New \succ Given information states among SO with bare argument with the number of such information states among SO strings where the object has a morphological presupposition trigger (which makes them exempt from the #New \succ Given constraint). As table 3 shows, there are 69 instances of such information state, which means that a non-given constituent preceding a given one is not a very rare information state in general.

Table 3: Rate of New \succ Given in finite clauses

	SO	S def/dem O
New \succ Given	0	69
Other (New \succ New, Given \succ New, Given \succ Given)	283	79

4 Simulating a mixed system: Rational Speech Act model

The Information Flow principle in (8) relates constituent order and morphological existential presupposition triggers as alternative markers of givenness in a type of a tradeoff relation. If a determiner is used, then the order of NPs does not matter for new/given encoding, and, conversely, if NPs meant to be interpreted as given precede NPs meant to be interpreted as new, determiners need not be used.

In terms of how they convey information structure, however, syntactic and morphological are not equivalent in so far as a constituent order NP₁ NP₂, incompatible with New \succ Given interpretation, is compatible with Given \succ New, Given \succ Given, and New \succ New interpretations, whereas a sequence NP₁ DET NP₂ is compatible with fewer information states: New \succ Given and Given \succ Given. Assuming that language users are aware of this, we can try to simulate the use of the two types of markers in a mixed system and compare the results with the quantitative data from diachronic French.

To do the simulations we use Rational Speech Act model (RSA, Frank & Goodman (2012)). RSA assumes Bayesian reasoning on the part of the speech act participants. Specifically, the beliefs of the Speaker and Listener are represented as probabilities they associate with different states of affairs. Probabilities that the Listener has before an act of communication are called prior probabilities. An utterance used in an act of communication is considered to be data that allow the Listener to update their knowledge by inferring posterior probabilities of the states of affairs. Interpretation (or probability update) at the so called literal listener level, is based solely on the literal meaning of the utterances (a pre-set relation between utterances and states of affairs). Then at the so called pragmatic speaker level the model takes into account the properties of the literal listener and an assumption that the speaker wants to maximize their chances to be understood (for the listener to infer from the utterances the state of affairs that the speaker means). It is this level that we use in our model of interaction of constituent order and morphological presupposition triggers.

In our simulation, we assume the states of affairs as in table 4 and possible utterances and correspondences between the two (literal meanings) as in table 5. We use DET here as a cover label for definite, demonstrative, possessive, and partitive determiners.

Table 4: States in RSA simulation

STATE	Given \succ New	New \succ Given	Given \succ Given	New \succ New
-------	-------------------	-------------------	---------------------	-----------------

Table 5: Literal meaning: utterances and corresponding information states

UTTERANCE	INFORMATION STATE
“DET O DET S”	Given \succ Given
“DET O S”	Given \succ Given, Given \succ New
“DET S DET O”	Given \succ Given
“DET S O”	Given \succ Given, Given \succ New
“O S”	Given \succ Given, Given \succ New, New \succ New
“S DET O”	Given \succ Given, Given \succ New, New \succ New
“S O”	Given \succ Given, Given \succ New, New \succ New

For the moment we assume that a priori all states of affairs and all utterances are equally likely. This means that, for instance, upon hearing “S O” a Literal Listener will end up with a uniform probability distribution over the states Given

\succ Given, Given \succ New, and New \succ New, as illustrated in figure 3.

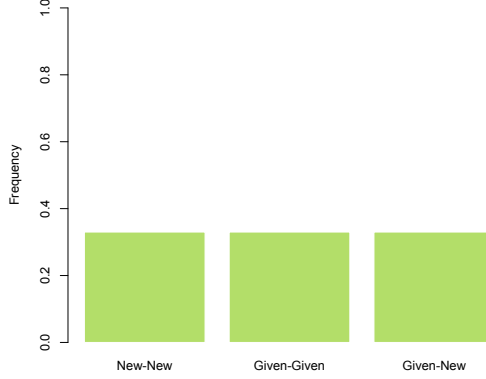


Figure 3: Literal Listener with uniform priors interpreting “S O”

A Pragmatic Speaker model generates inferences about what constituent orders a speaker is likely to use in order to convey a certain target information state given the assumptions of the Literal Listener and with the goal of maximizing the chances for the target state to be recovered. Figures 4–6 illustrate the inferences of the Pragmatic Speaker with regard to the Given \succ Given, Given \succ New, and New \succ New states, respectively. For instance, the model predicts that in order to convey Given \succ Given a speaker is most likely to use the order “DET S DET O” or “DET O DET S”. If we look again at table 5, we will see that among all the seven configurations eligible to convey a given Given \succ Given information flow, these two are the least ambiguous in that they are associated with only one information state.

*Givenness marking in a mixed system:
Constituent order vs. determiners*

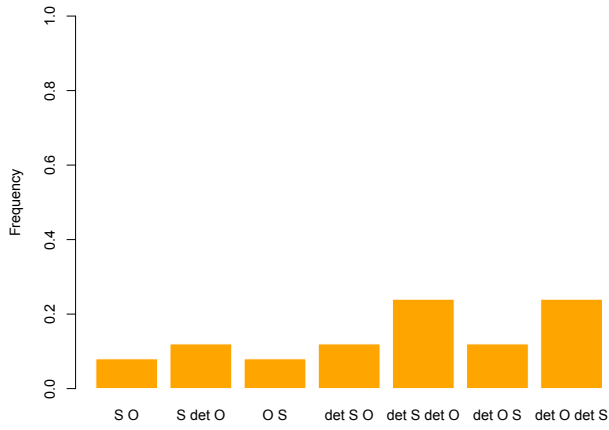


Figure 4: Pragmatic Speaker with uniform priors conveying Given \succ Given

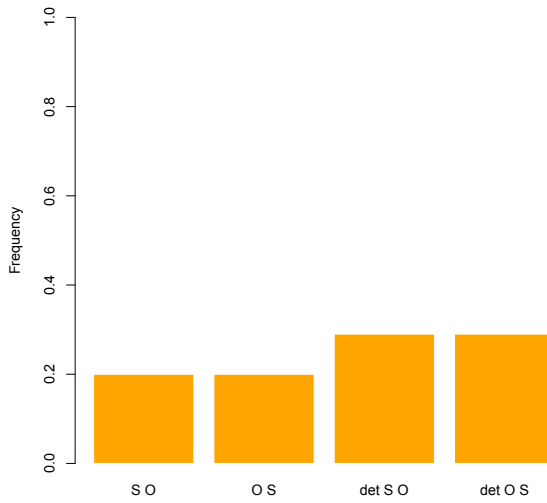


Figure 5: Pragmatic Speaker with uniform priors conveying Given \succ New

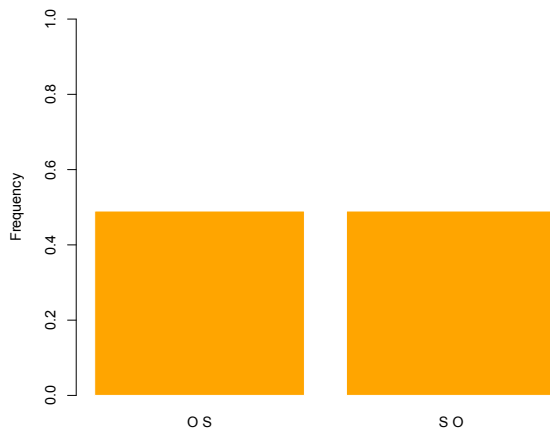


Figure 6: Pragmatic Speaker with uniform priors conveying New \succ New

The New \succ Given state can only be conveyed by one configuration, S_{DET} O (O_{DET} S order is not attested in the corpus and therefore is not part of our model).

A model of a Pragmatic Listener involves inferences with respect to the performance of a Pragmatic Speaker. That is, given a particular constituent order, this model makes inferences about most likely interpretations. Inferences for the “S O” configuration are shown in figure 7 .

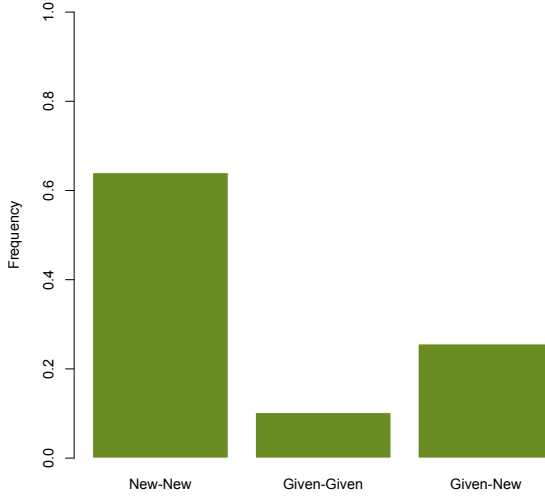


Figure 7: Pragmatic Listener with uniform priors interpreting “S O”

We see that a Pragmatic Listener model predicts that “S O” configuration is most likely interpreted as conveying a $\text{New} \succ \text{New}$ information state. As figures 4–6 show, to convey $\text{Given} \succ \text{Given}$ or $\text{Given} \succ \text{New}$, there are better candidates than “S O”, namely, “DET S DET O” or “DET O DET S” and “DET S O” or “DET O S”, respectively. That the model predicts “S O” to be most likely interpreted as $\text{New} \succ \text{New}$ corresponds to our intuition that the listener expects the speaker to use “DET S DET O” or “DET O DET S” and “DET S O” or “DET O S” for conveying the two other possible states. Since the Speaker visibly did not use either of those, the most likely interpretation is $\text{New} \succ \text{New}$, for which there is no better option than “S O”.

Let us now see how the Pragmatic Listener model fares compared to the historical French data. We classified all bare noun “S O” configurations in the corpus as $\text{New} \succ \text{New}$, $\text{New} \succ \text{Given}$ or $\text{Given} \succ \text{New}$ (recall that we did not find any bare noun “S O” corresponding to $\text{New} \succ \text{Given}$ information state). Figure 8 shows the distribution of information states among “S O” sequences in the corpus.

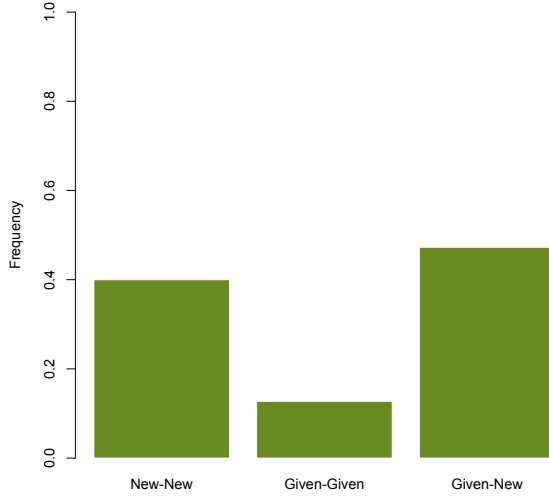


Figure 8: Distribution of informations states among “S O” in French, X-XVI c.

Comparing these results with the predictions of our RSA Pragmatic Listener model in figure 7, we see that in the actual data the frequency of $\text{Given} \succ \text{New}$ is higher than predicted, while the frequency of $\text{New} \succ \text{New}$ is lower.

Now, in our model we assumed that apriori all information states are equally likely (they had uniform priors). This is, however, most likely not the case (see Birner (2012) for references). We therefore need to make our information state priors more realistic. In order to do that, we used data from a syntactically annotated subcorpus of the Russian National Corpus, *Russian National Corpus* (2019). We classified 430 Russian transitive sentences with bare (i.e. without demonstrative or possessive determiners) nominal arguments (both “S O” and “O S”) according to their information state. The obtained distribution is plotted in figure 9.

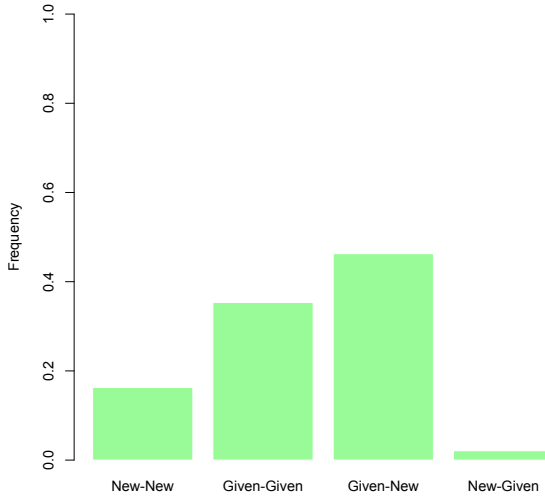


Figure 9: Distribution of informations states among transitive clauses with bare nominal arguments in the Russian National Corpus

We used these frequencies to set the priors for the information states in our RSA model. That is, instead of assuming that information states $\text{Given} \succ \text{Given}$, $\text{Given} \succ \text{New}$, $\text{New} \succ \text{Given}$, and $\text{New} \succ \text{New}$ are equally likely, we set their probabilities to 0.35, 0.46, 0.02, and 0.16, respectively. We then reran our Pragmatic Listener model, which now generates inferences for interpreting “S O” configuration as in figure 10, where it is plotted against the historical French data.

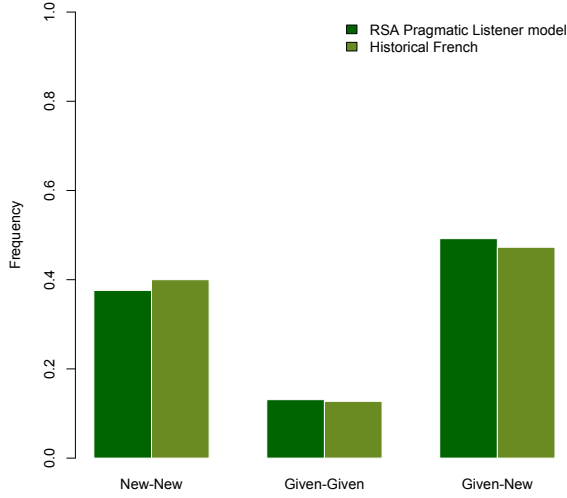


Figure 10: Distribution of informations states among “S O” as predicted by Pragmatic Listener model and in historical French corpus

As the figure shows, the shapes of the two distributions are remarkably similar, which means that our RSA Pragmatic Listener is a successful simulation of the pragmatic reasoning behind the historical French data.

The core assumptions of the simulation are encoded in the morphological entries in table 5, where the strings with presupposition-triggering determiners are less ambiguous than strings with bare NPs only and where New \succ Given information state cannot be conveyed by utterances involving bare nouns. This has the effect of predicting that, first, whenever there is a choice, speakers will be more likely to use strings with determiners than strings without, as this maximizes their chance to be understood (see figures 4 and 5), and, second, that pragmatically reasoning listeners will tend to interpret bare nouns as conveying information states which could not have been conveyed using presupposition-triggering determiners, such as New \succ New (see figure 7). The simulation results match historical French data very closely, while they contrast with the data we took from the Russian National Corpus where Given \succ Given is the second frequent information state of a transitive clause with bare nouns (see figure 9). We suggest that the differences is due precisely to the lack of definite determiners in Russian, which means that there is no better alternative for conveying Given \succ

Given than bare NPs, while in French “DET S DET O” is the best option (see figure 4).

5 Givenness marking and constituent order frequencies

In this section, we explore a connection between givenness marking and constituent order frequencies in historical French. Let us take another look at the constituent order distribution in table 3. Orders involving $O \succ S$ are markedly more rare than those with $S \succ O$, with one exception, namely, the OVS configuration. First, we suggest that the rarity of $O \succ S$ in medieval French, and thus the rarity of OSV and VOS, is a consequence of $\#New \succ Given$ on the assumption that subjects denote given information more frequently than objects. This assumption can be tested, at least at a first approximation, by looking at the distribution of determiners with subjects and objects. The rates of definite and possessive determiners and demonstratives with subjects and objects will be indicative of their respective tendencies to be associated with existential presupposition. Figure 11 shows the determiner distribution with subjects and objects per century.

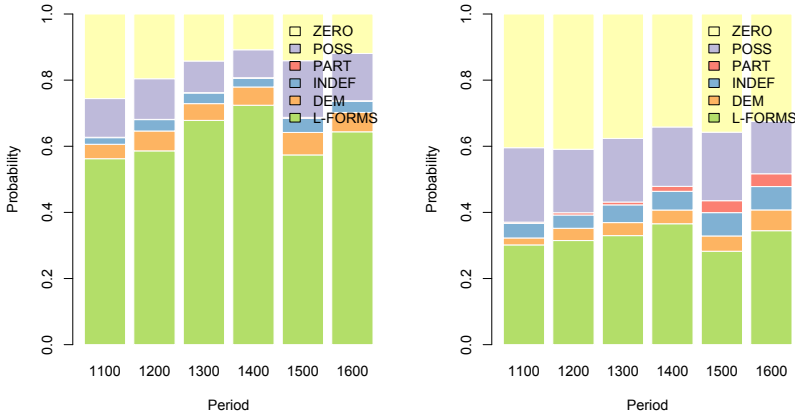


Figure 11: Determiner distribution: subjects vs. objects

Based on this approximation, we can estimate that during all periods subjects are at least 2–3 times more likely than objects to occur with a definite, possessive or demonstrative determiner, which indicates that subjects are much more

likely to satisfy the conditions on the use of presupposition-triggering determiners, namely, to denote an individual whose existence is entailed by the Common Ground.

Extrapolating this conclusion onto clauses with bare arguments, we expect that subject noun phrases denote properties whose extension is entailed by the Common Ground to be non-empty much more frequently than objects. This, in turn, means that the order $S \succ O$ is expected to align with the (licit) information state $\text{Given} \succ \text{New}$ much more frequently than the order $O \succ S$. We suggest that this is at least in part responsible for the very low frequency of OSV and VOS orders in historical French.³¹

We also observe that OVS order is more frequent than OSV and VOS. We suggest that OVS corresponds to a configuration of topic (situation) shift, where the preverbal position is associated with prosodic prominence. To probe into the properties of OVS, in figure 12 we plotted distributions of determiners in the object position in finite clauses with different constituent order. We take all the clauses with nominal objects and any type of subject (i.e. either nominal or pronominal or null).

³¹ According to Dryer (2013), in a sample of 1188 languages where a dominant constituent order can be established, there are only 40 languages (or $\approx 3\%$) where the dominant order involves $O \succ S$.

*Givenness marking in a mixed system:
Constituent order vs. determiners*

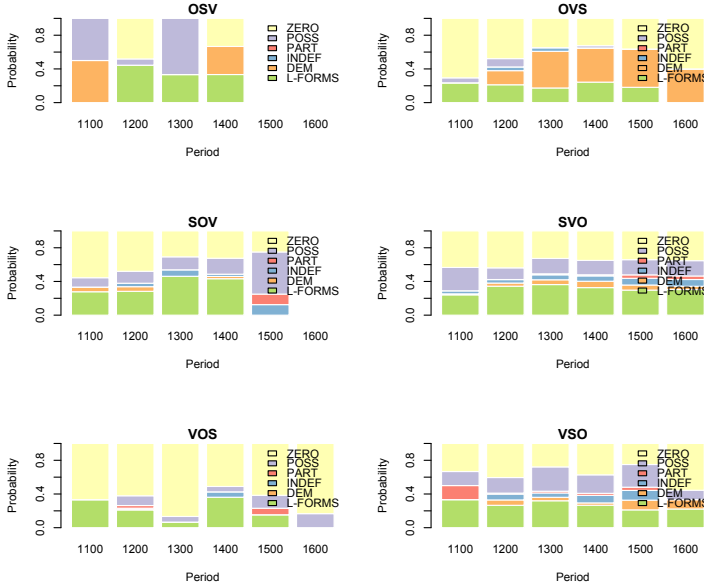


Figure 12: Determiner distribution in object position

Excluding from consideration numerically marginal (see table 3) and therefore highly erratic OSV and VOS patterns, we observe a similarity between object determiner distributions in SOV, SVO, and VSO configurations.

OVS stands out by an exceptionally high proportion of demonstratives in object position. To understand what that means for the status of OVS, let us consider the role of demonstratives in the information structure in modern languages.

The most notable feature of demonstratives is the requirement to have an antecedent (or an element in the extralinguistic reality serving as a referent). This property has been captured by assuming a silent individual pronoun in the structure of demonstrative phrases (Nunberg 1993, Elbourne 2008). On the view that pronominals are variables which get their value based on a context-determined mapping, in order for such structure to be interpretable, the context must involve a salient individual to which the assignment function will map the pronominal index.

Another potentially relevant fact is that the antecedent of a demonstrative is normally available in the immediately preceding context. According to Zulaica-Hernández & Gutiérrez-Rexach (2011: 180), in Spanish, 80% of demonstratives

have their antecedents in the immediately preceding utterance. Stevens & Light (2013: 204) report that in English 78.08% of demonstratives have antecedents that are discourse-new in the context immediately preceding the relevant demonstrative.

In addition, demonstratives, in contrast to definite determiners, are characterized by the requirement that the nominal predicate do not denote a singleton (relative to a certain domain, **Corblin:1987**). This is illustrated by the infelicity of (38) and (39) with demonstratives in the contexts implying uniqueness and by their felicity in contexts involving more than one individual with the relevant nominal property, as in (40).

(38) I fed #that/the dog. (If the speaker owns just one dog.)

(39) I saw #that/the brightest star.

(40) A woman_i entered from stage left. Another woman_j entered from stage right. That/#the woman_j was carrying a basket of flowers. (From Roberts 2002 & Wolter 2006: 74)

These three facts mean that an object noun phrase with a demonstrative requires an immediately preceding antecedent and that it also requires that the extension of the nominal predicate in the relevant situation do not correspond to a unique entity. An antecedent for a demonstrative must introduce a new entity, since an entity which had been introduced before would normally be realized as a pronoun or a noun phrase with a definite determiner, which is incompatible with the non-uniqueness requirement. In this respect, consider examples in (41) and (42).

(41) Workers painted a house_{new}. That house really needed it.

(42) Workers painted the house/it_{old}. #That house really needed it.

Based on such considerations, Bosch et al. (2003) formulate a Complementary Hypothesis, which states that personal pronouns pick up discourse topics as referents, while demonstratives prefer non-topical referents. Furthermore, Zulaica-Hernández & Gutiérrez-Rexach (2011: 175) argues for Spanish that “speakers use demonstratives to mark topic or subtopic shifts”. We thus conclude that the high rate of demonstratives with objects in OVS configurations indicates that the preverbal position was frequently used to indicate a shift in topic situation.

6 Conclusions

In this paper we explored organization of the information in a clause in light of Kučerová's (2012) proposal that givenness, when not expressed by dedicated morphemes, is monotonically marked from left to right. We proposed an amended version of the constraint that involves a non-presuppositional existential inference as the definition of givenness and tested it on a historical corpus of French, since the historical stages of French have both syntactic and morphological means of marking givenness. Our results show that the principle is borne out in the French historical data: in a corpus of 1,5 million words, we did not find any cases of New \succ Given in clauses with bare arguments. The data also bear out the prediction that in case Given is marked morphologically, the left-to-right monotonicity requirement does not apply.

We also built the principle into a game-theoretic simulation of the use of constituent order and presupposition-triggering determiners to convey an information state. The results of our simulation come very close to the empirical historical French data, suggesting that it is viable component of a model of pragmatic language use.

Finally, we also showed that *New \succ Given may provide insight into relative frequencies of various constituent orders.

Abbreviations

Acknowledgements

References

- Birner, Betty. 2012. *Introduction to pragmatics*. Vol. 38. John Wiley & Sons.
- Bosch, Peter, Tom Rozario & Yufan Zhao. 2003. Demonstrative pronouns and personal pronouns. German *der* vs. *er*. In *In proceedings of the eacl 2003. budapest. workshop on the computational treatment of anaphora*, 61–68.
- Büring, Daniel. 2003. On D-trees, beans, and B-accents. *Linguistics and Philosophy* 26(5). 511–545.
- Coppock, Elizabeth & David Beaver. 2015. Definiteness and determinacy. *Linguistics and Philosophy* 38(5). 377–435.

- Dryer, Matthew S. 2013. Order of subject, object and verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://wals.info/chapter/81>.
- Elbourne, Paul. 2008. Demonstratives as individual concepts. *Linguistics and Philosophy* 31(4). 409–466.
- von Fintel, Kai. 1996. *Specific generics*. <http://mit.edu/fintel/fintel-1996-specific-generics.pdf>. Handout from a Linguistics Colloquium given at Rutgers University.
- Frank, Michael & Noah Goodman. 2012. Predicting pragmatic reasoning in language games. *Science* 336(6084). 998–998.
- Harris, Martin. 1978. *The evolution of French syntax: A comparative approach*. Longman.
- Heim, Irene. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. University of Massachusetts, Amherst dissertation.
- Heim, Irene. 1991. Articles and definiteness. In Arnim von Stechow & Dieter Wunderlich (eds.), *Semantics: An International Handbook of Contemporary Research*. Berlin: De Gruyter.
- Heim, Irene. 2011. Definiteness and indefiniteness. In Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds.), *Semantics: an international handbook of natural language meaning*. De Gruyter Mouton.
- Kaiser, Georg & Michael Zimmermann. 2011. On the decrease in subject-verb inversion in French declaratives. In Esther Rinke & Tanja Kupisch (eds.), *The development of grammar: language acquisition and diachronic change; in honour of Jürgen M. Meisel*, 355–382. Amsterdam: Benjamins.
- Kratzer, Angelika. 2017. Situations in natural language semantics. In Edward N. Zalta (ed.), *The stanford encyclopedia of philosophy*, Winter 2017. Metaphysics Research Lab, Stanford University.
- Kroch, Anthony & Beatrice Santorini. 2010. *Penn Supplement to the MCVF (Martineau et al.)*
- Kučerová, Ivona. 2012. Grammatical marking of givenness. *Natural Language Semantics* 20(1). 1–30.
- Labelle, Marie & Paul Hirschbühler. 2018. Topic and focus in Old French V1 and V2 structures. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 63(2). 264–287.
- Lee, Ik-Hwan. 1995. An analysis of generic expressions in situation semantics. In *Proceedings of the 10th pacific asia conference on language, information and computation*, 19–28.

- Marchello-Nizia, Christiane. 1995. *L'évolution du français. Ordre des mots, démonstratifs, accent tonique*. Paris: Armand Colin.
2010. Corpus MCVF annoté syntaxiquement, (2005–2010), dirigé par France Martineau, avec Paul Hirschbühler, Anthony Kroch et Yves Charles Morin.
- Martineau, France. 2008. Un corpus pour l'analyse de la variation et du changement linguistique. *Corpus* 7. 135–155.
- Nunberg, Geoffrey. 1993. Indexicality and deixis. *Linguistics and Philosophy* 16(1). 1–43.
- Rinke, Esther & Jürgen M. Meisel. 2009. Subject-inversion in Old French: Syntax and information structure. In G. A. Kaiser & E.-M. Remberger (eds.), *Proceedings of the Workshop "Null-subjects, expletives, and locatives in Romance"*, 93–130. Konstanz: Fachbereich Sprachwissenschaft, Universität Konstanz.
- Roberts, Craige. 1996. Information structure in discourse: towards an integrated formal theory of pragmatics. *OSU Working Papers in Linguistics* 49. 91–136.
- Roberts, Craige. 2002. Demonstratives as definites. In K. van Deemter & Roger Kibble (eds.), *Information sharing: Reference and presupposition in language generation and interpretation*, 89–196. Stanford: CSLI Press.
- Russian National Corpus*. 2019. <http://www.ruscorpora.ru/en/index.html>.
- Sauerland, Uli. 2005. Don't interpret focus! In *Proceedings of Sinn und Bedeutung*, vol. 9, 370–84.
- Schwarz, Florian. 2009. *Two Types of Definites in Natural Language*. University of Massachusetts, Amherst dissertation.
- Schwarzschild, Roger. 1999. GIVENness, AvoidF and other constraints on the placement of accent. *Natural Language Semantics* 7(2). 141–177.
- Šimík, Radek & Marta Wierzba. 2015. The role of givenness, presupposition, and prosody in Czech word order: An experimental study. *Semantics and Pragmatics* 8. 3–1.
- Simonenko, Alexandra & Anne Carlier. 2019. *Maximality and situation-sensitivity: The evolution of French possessives*. https://www.academia.edu/37867707/Maximality_and_situation_sensitivity_The_evolution_of_French_possessives. Slides for a talk given at the workshop *Sorting out Definiteness* at DGfS 2019. University of Bremen. 6–8/03/2019.
- Simonenko, Alexandra, Benoit Crabbé & Sophie Prévost. 2018. Effects of text form on grammatical changes in Medieval French. A treebank-based diachronic study. *Diachronica* 35(3). 394–429.
- Skårup, P. 1975. *Les premières zones de la proposition en ancien français: essai de syntaxe de position*. Akademisk Forlag.

Alexandra Simonenko
Anne Carlier

- Stevens, Jon & Caitlin Light. 2013. The pragmatics of direct object fronting in historical English. *University of Pennsylvania Working Papers in Linguistics* 19(1). 23.
- Titov, Elena. 2012. *Information structure of argument order alternations*. UCL (University College London) dissertation.
- Vennemann, Theo. 1974. Topics, subjects and word order. In J. M. Anderson & Charles Jones (eds.), *Historical linguistics: proceedings of the first international conference on historical linguistics*, vol. 1, 339–376. Amsterdam: North Holland.
- Wolter, Lynsey Kay. 2006. *That's that: the semantics and pragmatics of demonstrative noun phrases*. University of California, Santa Cruz dissertation.
- Zulaica-Hernández, Iker & Javier Gutiérrez-Rexach. 2011. On the Information Status of Antecedents: Referring Expressions Compared. In *Proceedings of the Workshop Beyond Semantics: Corpus-based Investigations of Pragmatic and Discourse Phenomena*, Göttingen, Germany, 23–25.