The Rise and Fall of Periphrastic *Do* in Affirmative Declaratives

Relja Vulanović
Department of Mathematical Sciences
Kent State University
Stark Campus
6000 Frank Ave. NW
Canton, OH 44720, USA

rvulanovic@stark.kent.edu
http://www.stark.kent.edu/~rvulanovic

1. INTRODUCTION

The main source of data for my talk is (Ellegård, 1953) (E). In Part 1, E discusses the origin of periphrastic *do* and gives support to the causative origin theory. In Part 2, entitled *The Regulation of the Use of Periphrastic Do*, E presents and analyzes a large amount of data collected from 127 texts. Table 1 below is part of Table 7 in E. It shows the data for affirmative and negative declarative sentences and for affirmative and negative questions (negative imperatives are omitted).

	Aff. Decl.		Neg. Decl.		Aff. Quest.		Neg. Quest.	
Period	*do*	*n*	*do*	*s*	*do*	*s*	*do*	*s*
1390–1400	6	45000	0	—	0	—	0	—
1400–1425	11	4600	0	177	0	10	2	15
1425–1475	121	45500	11	892	6	136	2	23
1475–1500	1059	59600	33	660	10	132	3	24
1500–1525	396	28600	47	558	41	140	46	32
1525–1535	494	18800	89	562	33	69	34	22
1535–1550	1564	19200	205	530	93	114	63	21
1550–1575	1360	14600	119	194	72	56	41	7
1575–1600	1142	18000	150	479	228	150	83	45
1600–1625	240	7900	102	176	406	181	89	6
1625–1650	212	7200	109	235	116	24	32	6
1650–1700	140	7900	126	148	164	43	48	4
Swift(1710)	5	2800	61	9	53	3	16	0

Table 1. *do* = the exact count of sentences of each type with periphrastic *do*
n = a sample-based estimate (not the exact count) of the number of **all** affirmative declarative sentences
s = the exact count of sentences of each type with the 'simple' construction (i.e. without periphrastic *do*)

The counts were obtained from 10 randomly chosen pages of each text. I will combine here the two first periods into one period because of the small number of texts in each and will exclude Swift(1710).

Examples of sentence types
– Is ther no morsel breed **that ye do keep**? (Chaucer, *Monk's Tale*, line 444)
– Christ **dyd not praye** for . . . (E 305:319:11)
– Dolores mortis **not touched** hym or pynched hym (E 305:277:13)

– . . . for I **know not** myne owne religion (E 346:13:24)
– . . . **why dyde thou refrayne** from Ire, **why shewed thou not** vengeaunce vpon that moost vngentyll creature? (E 305:133:25)
– **Toke ye** hym in the quenys chamber? (E 243:1174:6)
– **Why do we not spede** vs . . . ? (E 305:195:35)

2. THE LOGISTIC CURVE

The 'S' shape of the graphical representation of the data is illustrated below in Figure 1 for affirmative questions, where $p = do/(do + s)$ and t stands for time. The particular points in time are the midpoints of each period. I used the SPSS software package to produce all graphs. This graph also shows a logistic curve obtained by the Curve Fit module of SPSS.

Kroch (1989a) was first to use the logistic curve to model p as a function of t:

$$p = \frac{1}{1 + e^{-at-b}}, \quad a, b = \text{const.}$$

The curve can be fitted to the data of all types of sentences but affirmative declaratives.

The meaning of the logistic curve is conveyed by the differential equation it solves,

$$\frac{dp}{dt} = ap(1 - p),$$

which shows that the rate of change is directly proportional to the quantity that changes, but slows down when that quantity approaches a certain maximum (1 in this case). The curve also can be used to model inhibited population growth, learning process, and other linguistic changes.

Other papers relying on E are (Kroch, 1989b) (K), and (Ogura, 1993) (O). Only the beginning of the development of periphrastic *do* in affirmative declaratives is modeled in these papers.

3. FITTING AFFIRMATIVE DECLARATIVES

The logistic curve can be generalized to approach any maximum $m > 0$ and not only to increase $(a > 0)$ but also to decrease $(a < 0)$. The generalization is

$$p = \frac{m}{1 + e^{-at-b}}$$

and this solves

$$\frac{dp}{dt} = \frac{a}{m}p(m - p).$$

Set $m = .1$ and split the data into two groups: first seven points and last five points (year 1562.5 and $p = .093$ are the coordinates of the point shared by both groups). Figures 2 and 3 show the fit, which is relatively good: the coefficient of determination is $R^2 = .913$ for the rising data and $R^2 = .823$ for the falling ones. As a comparison, $R^2 = .911$ for the data in Figure 1, and $R^2 = .842$ and .761 for negative declaratives and negative questions respectively.

It can be concluded that if some model can be used to describe the rise of a (linguistic) quantity in one part of the change, then it also can be modified to describe its fall in another part, and therefore a combination of the two processes. This is not surprising in the context of affirmative declaratives, since the logistic curve can be used to model various syntactic changes, cf. K. The rise and fall of periphrastic *do* in affirmative declaratives can be viewed as two connected syntactic changes.

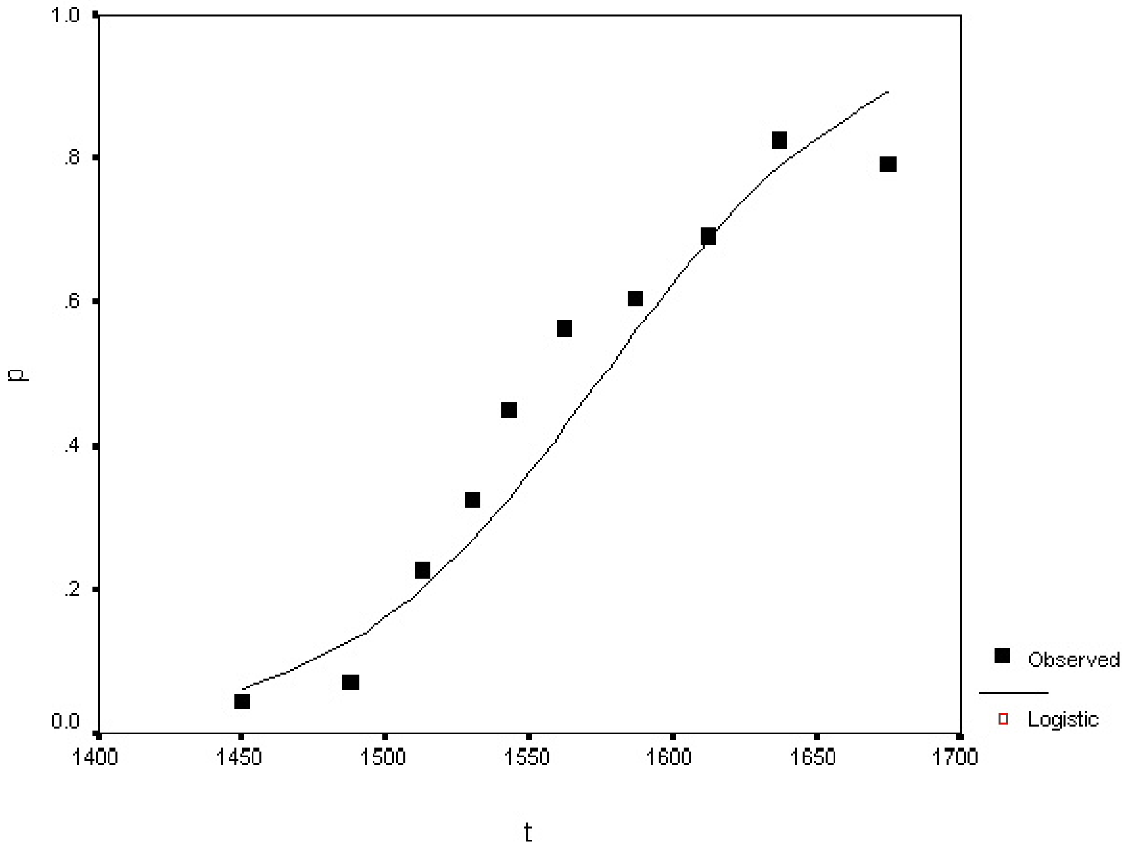

Fig. 1. Affirmative questions.

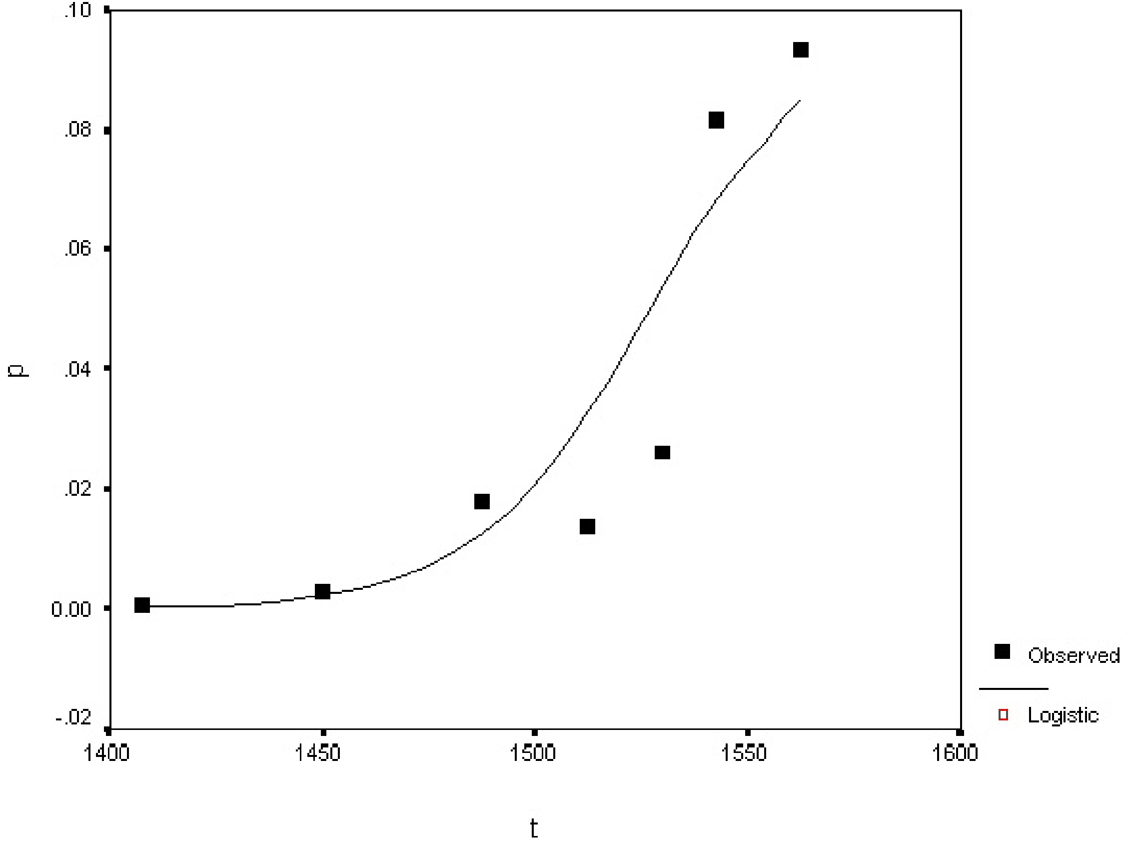

Fig. 2. Rise of 'do' in affirmative declaratives.

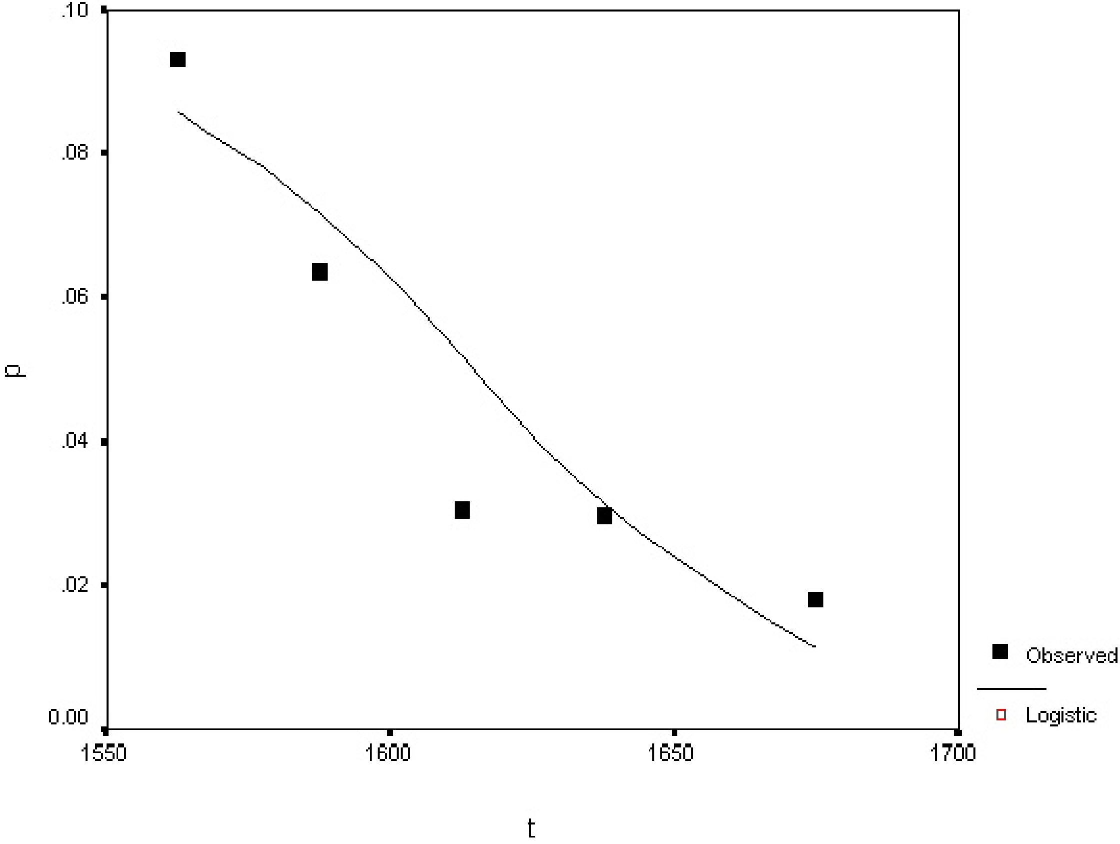

Fig. 3. Fall of 'do' in affirmative declaratives.

4. EXPLANATION OF THE CHANGE

K and O disagree on some points of the development of periphrastic *do* (in terms of the time line and interpretation), but they do[1] agree on the actuation scenario. In short, periphrastic *do* is motivated by the need to lexically support affixes:

1. the collapse of the subjunctive mood in ME, leads to the use of modals instead of the subjunctive

2. modals lose their status as main verbs and become auxiliaries appearing in the INFL(ection) or AUX(iliary) positions of the phrase structure tree

3. V-to-INFL raising is gradually lost in main verbs and the tense marking is performed through affix hopping – a transfer of the affix from INFL to the main verb in its deep structure position

4. affix hopping is blocked (eg. by the negator *not* or the subject between INFL and the main verb in negative declarative sentences or questions respectively)

5. periphrastic *do* is inserted to provide lexical support for the affixes in INFL

As for affirmative declaratives, K's view is that

- there is nothing to block affix hopping and the failure of V-to-INFL raising does not force the use of periphrastic *do*

- its frequency therefore never goes higher than 10%

- surface reflexes of V-to-INFL raising have to be reanalyzed grammatically after its loss (which K thinks happened after period 7, followed by independent development of *do* in different contexts)

- one of the results of this reanalysis is the loss of periphrastic *do* in affirmative declaratives

O disagrees:

- if K is right on the first point, then why is periphrastic *do* used in affirmative declaratives at all?

- initially, affix hopping and the insertion of periphrastic *do* were equally available regardless of whether there was an interfering element between INFL and the main verb or not

- adverbs preceding the main verb started blocking affix hopping

- the use of affix hopping got restricted to SV sentences, while the use of periphrastic *do* got restricted to SAdvV sentences

- this restricted use eventually won the competition against the unrestricted use of *do* and affix hopping [note that this is not fully explained]

Both K and O use some hypothetical examples ("mental experiments") to support their respective analyses. I would like to use here the grammar efficiency model to explore the role of emphatic *do* in the change and the possibility that periphrastic *do* was reanalyzed as emphatic *do*[2]. Adverbs will be excluded from consideration, as every modeling requires a simplification.

On p. 172, E gives a count of identifiable emphatic *do* in affirmative declaratives over 5 longer periods. Table 2 shows percentages which can be calculated from Tables 7 and 8 in E. I will use these data in the grammar efficiency model. I will assume that the reanalysis of periphrastic *do* as emphatic *do* starts in period II.

[1]This is an example of emphatic *do*.
[2]K and O do not discuss emphatic *do* at all.

Period		p = % all *do*	ε = % emph. *do*	% emph. *do* as part of all *do*
I:	1400–1500	1.09%	0.00%	0.42%
II:	1500–1525	1.38%	0.02%	1.26%
III:	1525–1550	5.42%	0.09%	1.65%
IV:	1550–1600	7.67%	0.06%	0.76%
V:	1600–1700	2.57%	0.09%	3.38%

Table 2. Periphrastic *do* and emphatic *do* in 5 periods.

5. GRAMMAR EFFICIENCY

The concept of grammar efficiency was introduced in (Vulanović, 1991) and further developed in (Vulanović, 1993, 1999, unpublished manuscript). It has been used to model various syntactic changes, see also (Vulanović, 1995, 1997). Grammar efficiency is defined by $\mathit{Eff} = \mathit{Info}/\mathit{Con}$, where *Info* is the number of types of information (mainly syntactic functions) conveyed by the grammar and *Con* is a measure (a positive real number) of the set of grammatical devices used by the grammar to convey the desired information. $\mathit{Con} = k + \alpha R$ with k being the number of grammatical conveyors (like word classes, grammatical categories, etc.). R is a measure of how much information is conveyed by word order which is a special kind of a conveyor. R is greater (and therefore *Eff* is less) if word order is less free or if the grammatical structure permits more ambiguous sentences. Finally, α is a weight (a positive real number) balancing the relative importance of grammatical conveyors and word order. Numerical values of *Info* and k are easily found, whereas R and α require special formulas based on combinatorics.

Some simple examples – SVO sentences

Set $\mathit{Info} = 3$ as 3 syntactic functions are to be conveyed: S, P(redicate) and O.

$k = 2$. N and V are used as grammatical conveyors. Three orders of S, P, and O (of the theoretically possible $3! = 6$ orders) can be permitted, so that there is no ambiguity (e.g. those in which S precedes O). Then,

$$R = \frac{3!}{3-0} - 1 = 1.$$

The zero above indicates the number of ambiguous sentences. The 1 subtracted from the fraction is part of the formula ensuring that $R = 0$ when word order conveys no information.

The above structure can be considered maximally efficient. For those, define $\mathit{Eff} = 1$, which gives

$$2 + \alpha \cdot 1 = 3 \Longrightarrow \alpha = 1.$$

Then, for instance, efficiency of English is

$$\mathit{Eff} = \frac{3}{2 + 1 \cdot \left(\frac{3!}{1} - 1\right)} = \frac{3}{7} = 0.429.$$

If all word orders were permitted, all sentences would be ambiguous, resulting in $\mathit{Eff} = 0$:

$$R = \frac{3!}{6-6} - 1 = \infty, \quad \mathit{Eff} = \frac{3}{\infty} = 0.$$

$k = 3$. Let grammatical conveyors be Nom(inative), Acc(usative), and V. Then all six permutations of S, P, and O can be permitted.

$$R = \frac{3!}{3!} - 1 = 0 \quad \text{(word order conveys nothing).}$$

Theoretical efficiency of Latin is

$$Eff = \frac{3}{3 + 1 \cdot 0} = 1 \quad \text{(maximum efficiency)}.$$

Latin with 79% SOV and 21% SVO sentences can be represented as follows:

$$R = \frac{3!}{1 + \frac{21}{79}} - 1 = 3.74,$$

$$Eff = \frac{3}{3 + 1 \cdot 3.74} = 0.445.$$

Application to Periphrastic *do* in Affirmative Declaratives

In a syntactic change, *Eff* is evaluated in each stage and then the change can be represented by a sequence of numbers or even graphically. When a new structure is introduced, *Eff* decreases and then gradually increases as the new structure is used more and more. Based on the data collected by E and my previous work, it can be expected that the efficiency graph of the change affecting periphrastic *do* in affirmative declaratives shows the decreasing-increasing pattern or a combination of such patterns.

Several models are considered below. 7 or 9 stages are used (the starting stage, the five stages of Table 2, and 1 or 3 finishing stages). Examples of the formulas are provided here for Model A only.

A. Model without emphatic ***do*** (Figure 4)
Info = 2 [S and P(redicate) are conveyed], $\alpha = 1$
Start. NV analyzed as SP. $R = \frac{2!}{1} - 1 = 1$, $Eff = \frac{2}{2+1 \cdot 1} = \frac{2}{3} = .667$
Stages I–V. NV, N[*do*V] (*do* and V treated as one unit – they together convey P)
$R = \frac{2+2}{1+\frac{p}{1-p}} - 1 = 3 - 4p$, $Eff = \frac{2}{2+p+1 \cdot (3-4p)} = \frac{2}{5-3p}$
End. Return to the initial structure, $Eff = \frac{2}{3}$.

B. Model with emphatic ***do*** (Figure 5)
Start–I. Like in A.
II–V. NV and N[*do*V] analyzed as SP, N*do*V analyzed as SEmP (Em=emphasis), (values of ε from Table 2 are used when calculating *Eff*)
End. NV analyzed as SP, N*do*V analyzed as SEmP

C. Models of a hypothetical development: the use of periphrastic *do* continues to increase, p approaching the value of 1 (but kept less than 1)
C.a. Model without emphatic ***do*** (Figure 6): stages Start–IV like in A.
C.b. Model with emphatic ***do*** (Figure 7): stages Start-IV like in B.

D. Comparison model: negative declaratives (Figure 8)
Start. NV*not*, N*not*V
I–VII. NV*not*, N*not*V, N[*do not*]V, the latter gradually increasing in use
End. N[*do not*]V

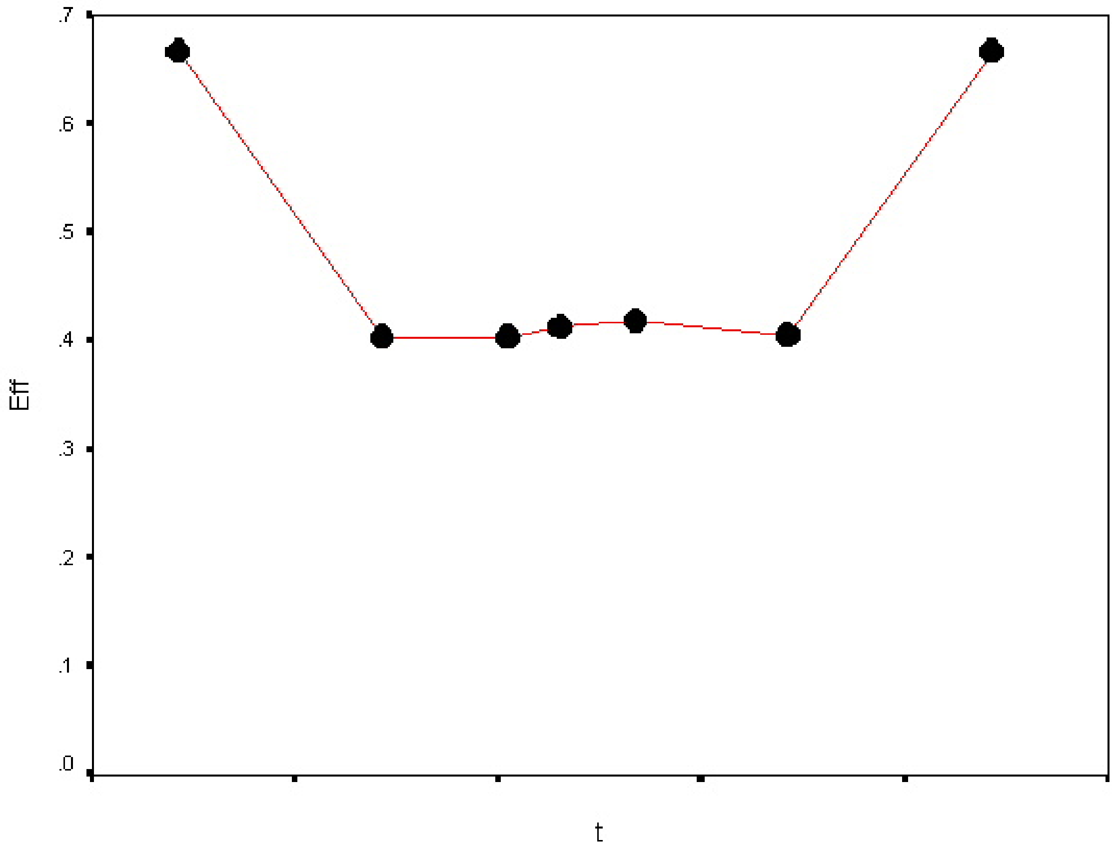

Fig. 4. Efficiency model without emphatic 'do'.

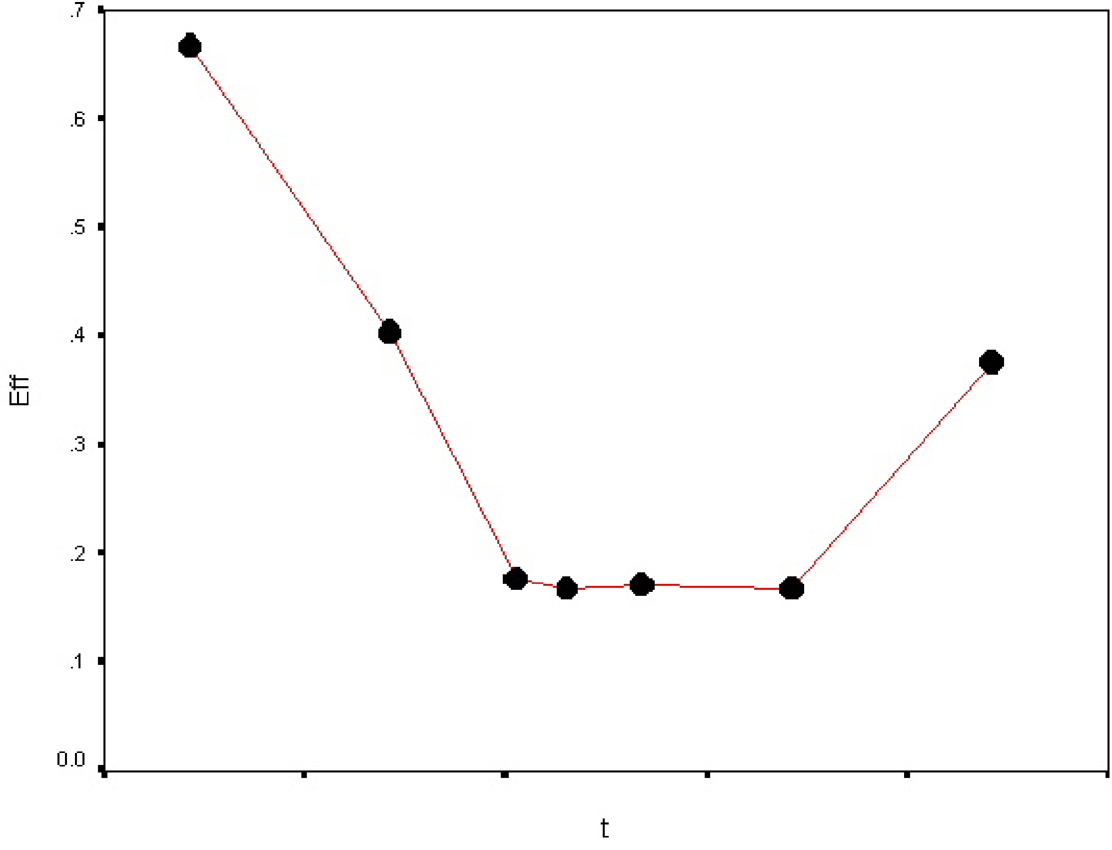

Fig. 5. Efficiency model with emphatic 'do'.

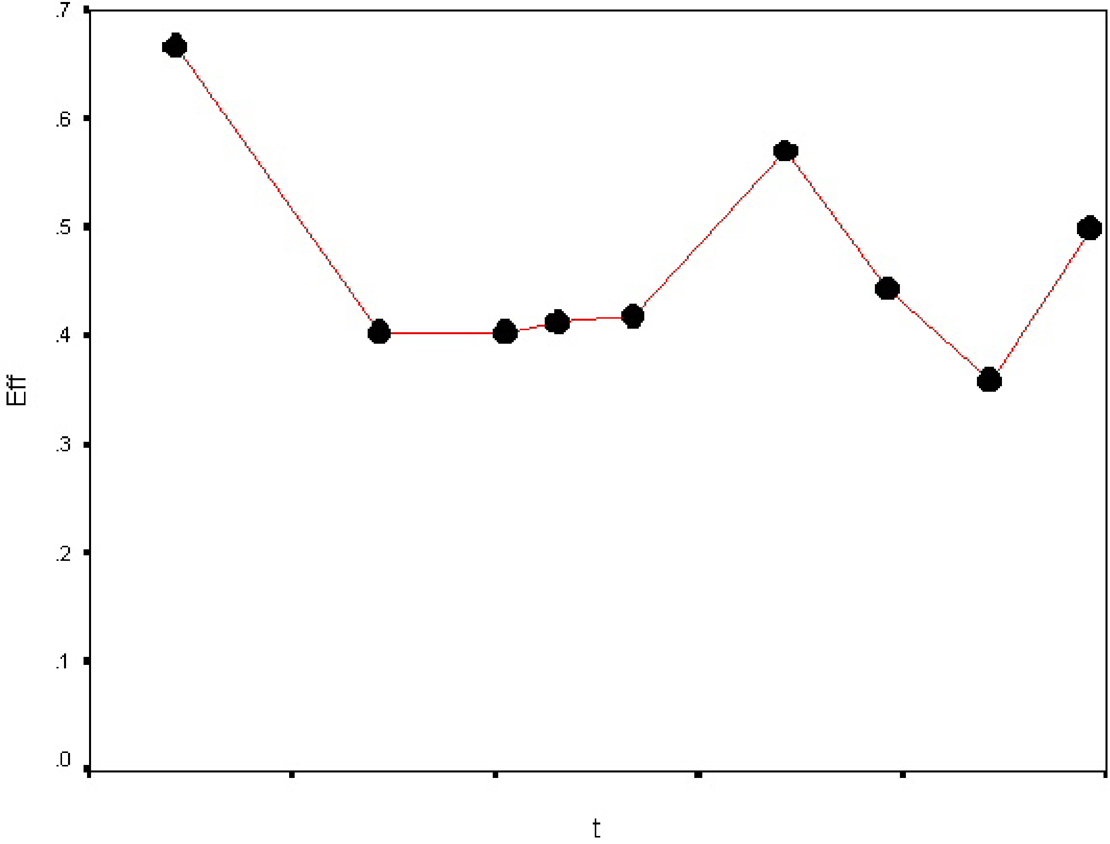

Fig. 6. Hypothetical rise of 'do' without emphatic 'do'.

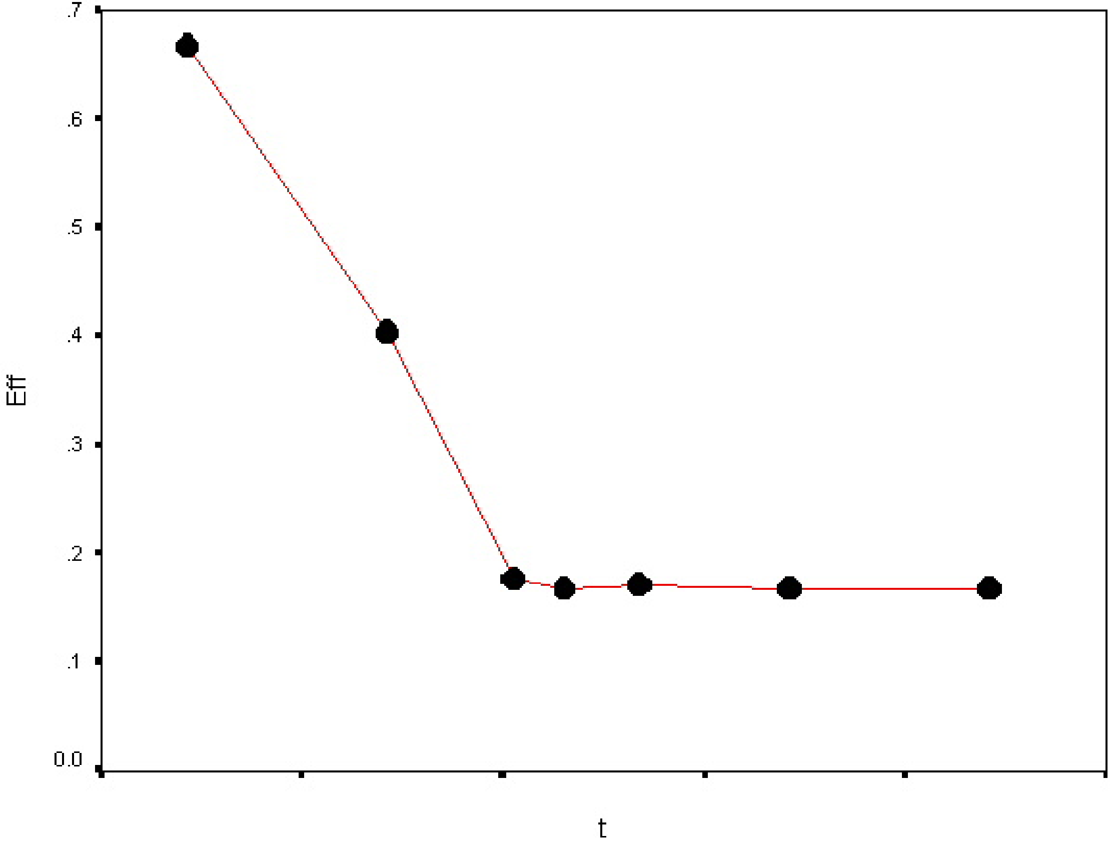

Fig. 7. Hypothetical rise of 'do' with emphatic 'do'.

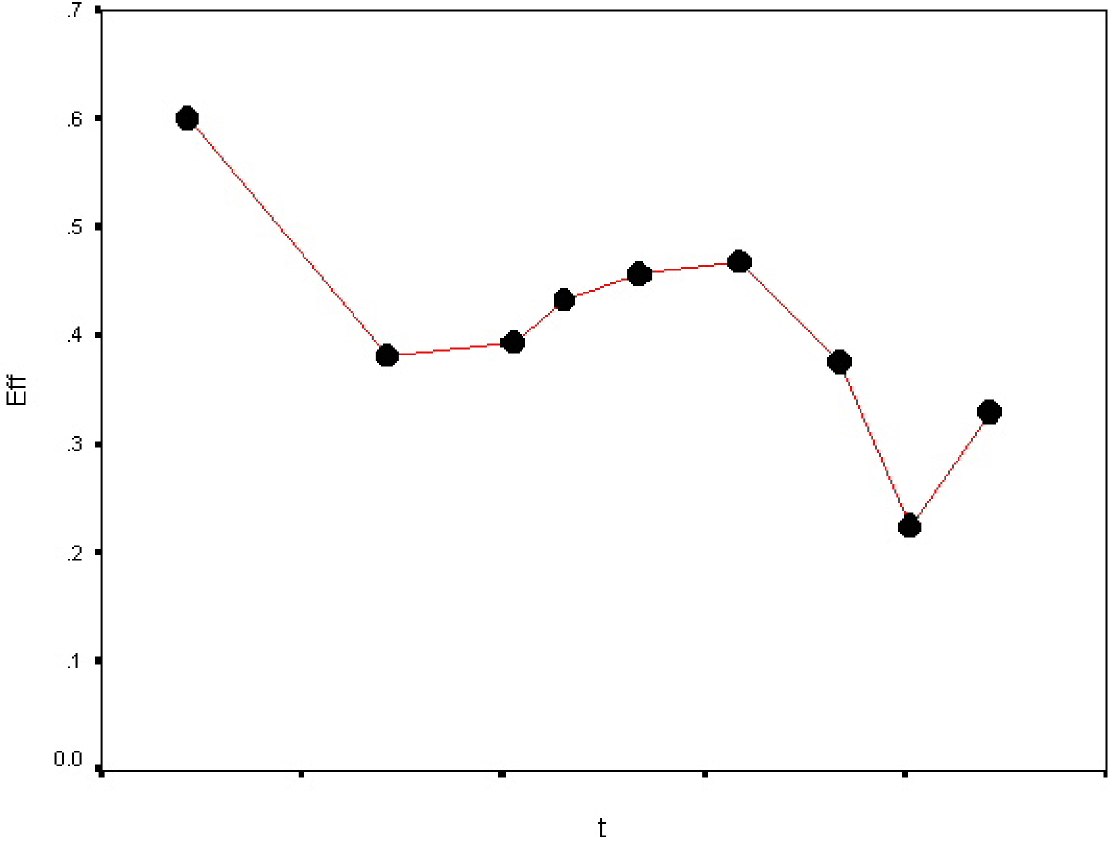

Fig. 8. Negative declaratives.

Conclusion

The hypothetical model C.b of continuous rise of *do* with emphatic *do* does not show the decreasing-increasing behavior present in all grammar efficiency graphs of syntactic changes I have modeled so far in my papers. It seems safe to assume that this kind of change is not very likely to happen. This confirms that periphrastic *do* cannot continue to rise in the presence of emphatic *do* because of the ever-present ambiguity.

The hypothetical model C.a of continuous rise of *do* without emphatic *do* still shows a possible development. Its efficiency graph is shaped similarly to that of negative declaratives. The reason why this development did not happen is emphatic *do*.

Model A and its efficiency graph show a peculiarity absent in all syntactic changes I have considered. The decreasing-increasing shape of the graph is present and I have had examples of changes finishing with the same efficiency as the starting one. However, in all such changes, the last grammatical structure is **different** from the first one, it just so *happens* that they have the same efficiency. In A, the starting and ending structures are exactly the same like the whole change was for nothing. Therefore, model B is more plausible than model A and emphatic *do* has to be considered a factor in the development of periphrastic *do* in affirmative declaratives.

A reanalysis of this nature could not happen in other sentence-types with periphrastic *do*. The different contexts are certainly connected in the change, but affirmative declaratives require a special analysis.

References

Alvar Ellegård, 1953. *The Auxiliary Do: The establishment and regulation of its use in English*. Stockholm: Almquist & Wiksell.

Anthony S. Kroch, 1989a. "Function and grammar in the history of English: Periphrastic *do*". pp. 133-172 in *Language Change and Variation*, R.W. Fasold and D. Schiffrin, eds. Amsterdam: John Benjamins.

Anthony S. Kroch, 1989b. "Reflexes of grammar in patterns of language change". *Language Variation and Change* **1**, 199–244.

Mieko Ogura, 1993. "The development of periphrastic *do* in English: A case of lexical diffusion in syntax". *Diachronica* **10**, 51–85.

Relja Vulanović, 1991. "On measuring grammar efficiency and redundancy". *Linguistic Analysis* 21:201–211.

Relja Vulanović, 1993. "Word order and grammar efficiency". *Theoretical Linguistics* 19:201–222.

Relja Vulanović, 1995. "Model–based measuring of syntactic change". *J. Quantitative Linguistics* 2:67–76.

Relja Vulanović, 1997. "The development of negation in French: A quantitative model". *J. Quantitative Linguistics* 4:276–280.

Relja Vulanović, 1999. "Grammar efficiency and the historical development of word order in French". pp. 193–206 in *Issues in Mathematical Linguistics*, C. Martín-Vide, ed. Amsterdam: John Benjamins.

Relja Vulanović, unpublished manuscript. "On grammar efficiency and its applications".