

(How) can causative constructions be predicted?

Gaëtanelle Gilquin¹, Éric Lecoutre²

¹ F.N.R.S. / CECL, LIGE, UCL – Place Blaise Pascal, 1 – 1348 Louvain-la-Neuve – Belgique
gilquin@lige.ucl.ac.be

² Institut de statistique, UCL – Voie du Roman Pays, 20 – 1348 Louvain-la-Neuve – Belgique
lecoutre@stat.ucl.ac.be

Abstract

This study investigates whether English causative constructions can be predicted, and if so, how. The technique used here is that of the decision tree, which emerges as the most powerful tool for our purposes. Some results are given and it is shown how these results can be applied, not only to a scientific description of causative constructions, but also, provided some changes are implemented, to the field of second or foreign language acquisition.

Résumé

Le but de cette étude est de déterminer si les constructions causatives anglaises peuvent être prédites, et si oui, comment. La technique utilisée à cet effet est l'arbre de décision, qui apparaît comme l'outil le plus approprié pour ce faire. On donnera quelques résultats, et on montrera que ces résultats peuvent être appliqués, non seulement à une description scientifique des constructions causatives, mais aussi, à condition que certains changements soient effectués, au domaine de l'acquisition d'une langue seconde ou étrangère.

Keywords: decision tree, prediction, causative constructions.

1. Introduction

As pointed out by Kemmer and Verhagen (1994:115), “[t]he grammar of causative constructions has inspired what is probably one of the most extensive literatures in modern Linguistics.” This they explain by the “fascinating complexity of causatives both within particular languages and cross-linguistically,” as well as the “tacit recognition by many linguists that an understanding of causatives is fundamental to an understanding of clause structure as a whole.” Despite this complexity and centrality, and although causative constructions have been approached from a wide variety of perspectives – including (but not limited to) Generative Semantics and its classic derivation of *kill* from CAUSE BECOME NOT ALIVE (cf. McCawley, 1968), the universal-typological perspective (e.g. Comrie, 1976 or Wierzbicka, 1998) or the approach of functional grammar (see Dik, 1980) – it must be recognised with Altenberg and Granger (2001: 184) that “[i]t is very difficult to find a good description of the usage differences between [causative] verbs.” Patchiness, focus on formal, rather than semantic aspects, and general lack of reliability are among the weaknesses of the descriptions of causative constructions found in the literature. This can be linked to the lack of empirical foundations of such descriptions for, as rightly emphasised by Kemmer and Verhagen (1994: 148), “a complete understanding of causatives must (...) take into account *empirically attested* semantic patternings (...), an area which has been relatively neglected in the literature so far” (emphasis added). The aim of the study presented here is precisely to take such empirically attested patternings into account, and see whether the choice of a particular causative verb can be predicted on this basis.

2. Material

Four English causative verbs have been investigated, viz. *cause*, *get*, *have* and *make*, as they are used in so-called periphrastic causative constructions (e.g. *He made her leave the room* or *I had the car repaired*). The data come from the British National Corpus (BNC) and consist of some 5 million words of spoken English (spontaneous conversations and broadcast discussions) and 5 million words of written English (academic prose from various fields). The number of causative constructions retrieved is shown in Table 1.

	n	%	/ 100,000 words
<i>CAUSE</i>	200	5.6%	2.04
<i>GET</i>	1,310	36.7%	13.36
<i>HAVE</i>	813	22.7%	8.29
<i>MAKE</i>	1,251	35.0%	12.76
TOTAL	3,574	100.0%	36.46

Table 1. Raw frequency (n), percentage and relative frequency per 100,000 words of the four causative verbs in the corpus

For each causative construction, a number of (syntactic and semantic) variables were examined and stored in a database, such as the tense of the verb, the animate or inanimate nature of the CAUSER,¹ or the form of the EFFECT (bare infinitive, *to*-infinitive, present participle or past participle). In total, some 40 variables were thus analysed.

3. Methodology

Several methods can be applied in order to predict the use of a particular word or construction, but they are not all equally appropriate to deal with causative verbs. Without any other available information than the observed frequencies, one can envisage two different ways to predict the causative verb, viz. raw prediction and naive prediction. The raw prediction always returns the same output, namely the most frequent verb, *get*, which leads to a prediction rate of 36.7% (i.e. the observed proportion of causative constructions that use *get*). The naive prediction consists in returning a random verb with a probability corresponding to its empirical proportion. The naive predictor results in a global rate of prediction of 30%.² Not only are these two methods simplistic, however, but they do not provide any explanation for why one form should be preferred rather than another. To provide such an explanation, one should investigate variables that might account for the choice of a particular causative. As can be expected from Kemmer and Verhagen's remark on the complexity of causatives, pinpointing a single variable will not do. Thus, Guierre's (1959: 126-7) claim that the choice between the different causatives can basically be reduced to a distinction between active sense ("acting," e.g. *She made her brother read the book*) and passive sense ("undergoing," e.g. *He had his watch fixed*) is obviously inaccurate, as most verbs can be used with both an active and passive meaning (compare *She made her brother read the book* with *She made her views*

¹ The CAUSER refers to the initiator of the causative process (e.g. *He had his house built in 1980*). The other terms used to designate the different elements of the causative construction are the CAUSEE, the entity which is changed or influenced by the CAUSER and carries out the EFFECT (e.g. *The teacher made him read the book*). The EFFECT represents the action thus performed (e.g. *He got the video working*) and can be followed by an object, called the PATIENT (e.g. *The drought has caused millions of people to leave their homes*).

² The formula to calculate this is: $\sum (p_i)^2$, where p_i = empirical proportion in the corpus of the i^{th} form.

known, or *He had his watch fixed* with *He had his class write an essay*). So clearly, several variables have to be taken into account. One way of doing this is by means of discriminant function analysis. This technique is used to determine which variables discriminate between a number of target categories (here, the categories of *cause-*, *get-*, *have-* and *make-* constructions). The variables thus emerging are the best predictors for the category. Although the technique can be used with excellent predictive power (cf. Gries's (2003) analysis of the dative alternation), it is important to bear in mind that it judges the variables individually, not in combination with one another. That this difference can be crucial appears from the following example. In our data, the CAUSER and CAUSEE with *make* are predominantly animate (although the former with a small majority), while the EFFECT tends to be non-volitional. Yet, the most frequent combination is not animate CAUSER + animate CAUSEE + non-volitional EFFECT, but **inanimate** CAUSER + animate CAUSEE + non-volitional EFFECT (with a proportion of 28.40%, against 17.84% for the former combination). In other words, it turns out that the combination of the most frequent individual variables does not perforce correspond to the most frequent combination of variables. There are at least two techniques that examine the different variables in relation to one another, viz. neural networks and decision trees (see Berry and Linoff, 1997). Since decision trees, unlike neural networks, are directly expressed as understandable rules, they should be preferred in an analysis like ours which aims not only to predict the choice of one verb or another, but also (and above all) to explain this choice. The technique of decision trees has been applied successfully by Duhoux and Lecoutre (2003), who sought to establish the variables that influence the choice of verbal aspect in ancient Greek and managed to predict over 60% of the forms.

Decision trees can be built in the SAS (Statistical Analysis Software) System – “Enterprise Miner”. At the root of the tree is the whole set of the training data. By means of the χ^2 test, all the variables are examined in order to determine the best candidate for a split, that is, the candidate that will divide the training data into two subgroups with the best predictive results. The same process is applied to the nodes thus obtained and is repeated until we reach a terminal node, called a “leaf”. Each leaf corresponds to a particular prediction and a precise rule, which can be retrieved by looking at the path that connects the root to the leaf. The rules take the following form: “If $X = a$ and $Y = b$, then Verb = v ,” followed by the percentage that v represents in the leaf, as well as the number of records that are concerned by the rule. The predictive power of the tree can be assessed by calculating the number of records that are correctly classified according to the tree. The reliability of the tree can then be checked by applying it to previously unseen data. If the tree performs as well with the new data as with the training data, the tree can be said to be reliable. Here, the sample has been divided into 60% of training data and 40% of validation data.

4. Results

Despite the caveat mentioned with respect to discriminant function analysis, this technique was tested against our data. While the verb *get* reaches a very good 88% of correctly predicted cases, the proportion falls to 67.5% with *have*, 36.5% with *cause* and under 25% with *make*. As a comparison, the decision tree with the same variables has a total predictive power of some 83%, taking eight leaves into account.³

However, an analysis of the results of the decision tree shows that it might be preferable to make a selection among the variables, rather than using them all, and in fact, to change the

³ Beyond eight leaves, the results of the training data start to diverge from the validation data.

nature of the targets altogether. With no prior selection, the variable that best discriminates among the target classes is the form of the EFFECT, with the first rule stating that “If Form = Infinitive, then Verb = MAKE (93%).” This rule, which predicts 487 occurrences of *make* (that is, 90% of the total occurrences of *make* in the training data), tells us little more than an analysis of the frequency of the different complements with *make*. Moreover, this variable can be said to partially overlap with the target, since causative *make* is only possible with certain types of complements, including bare infinitive. Generally speaking, we do not primarily choose the verb *make* because we want to put the EFFECT in the bare infinitive, but we choose the bare infinitive because it is among the types of complements that are acceptable after *make*. Put differently, we tend to choose the causative verb first (according to rules that will have to be discovered), and then select one of the possible complements.⁴ It was therefore decided to include the complement in the target and so turn the four targets into ten targets, that is each verb with the different complements it accepts, viz. CAUSETO (*cause* + *to*-infinitive), GETTO (*get* + *to*-infinitive), GETPP (*get* + past participle), GETPRP (*get* + present participle), HAVEINF (*have* + infinitive), HAVEPP (*have* + past participle), HAVEPRP (*have* + present participle), MAKEINF (*make* + infinitive), MAKETO (passive *make* + *to*-infinitive) and MAKEPP (*make* + past participle) (see Table 2 for the frequency of the different patterns).⁵ This decision is supported by the analysis of the individual variables, which underlines the specificity of the different causative structures (e.g. large proportion of inanimate CAUSEES in present participle constructions with *get* and *have*, but not in infinitive constructions).

	n	%	/ 100,000 words
CAUSETO	200	5.6%	2.04
GETTO	366	10.2%	3.73
GETPRP	129	3.6%	1.32
GETPP	815	22.8%	8.31
HAVEINF	72	2.0%	0.73
HAVEPRP	70	2.0%	0.71
HAVEPP	671	18.8%	6.84
MAKEINF	1,120	31.3%	11.42
MAKETO	100	2.8%	1.02
MAKEPP	31	0.9%	0.32
TOTAL	3,574	100.0%	36.46

Table 2. Raw frequency (n), percentage and relative frequency per 100,000 words of the ten causative patterns in the corpus

⁴ This is a simplification, to some extent, as some types of complements are not possible with all four causatives (cf. present participle, possible with *get* and *have*, but not with *cause* and *make*) and so might contribute to the choice of, say, one pair of verbs rather than another. However, the definition of the new targets will not ignore the parameter of the form of the EFFECT.

⁵ Other types of complements are sometimes found in the corpus data, cf. (i) or (ii), but these are very marginal and so will not be taken into account here.

- (i) What's **made** you *to think* of that? <BNC:S:KBB 612>
- (ii) How do we **get** them *drink* more? <BNC:S:KBD 4671>

In addition, a number of variables were removed,⁶ either because they had become redundant with the introduction of the new targets (e.g. voice of the EFFECT), or because they had not emerged in the decision tree built on the basis of all the variables (this is the case of the PATIENT, which, beyond its mere presence or absence, does not seem to influence the choice of the causative).⁷ The remaining variables represent a total of 15.

The new tree has a total predictive power of some 73%. Granted, this is less good than the first decision tree, but still much better than a naive prediction (the naive predictor, it will be reminded, has a power of 30%). And in this new tree, there is no overlap between the targets and the variables. It is important to bear in mind, however, that not all targets are equally well predicted by the tree. In fact, there are two structures that are not predicted at all (at least at an “acceptable” level, i.e. at a level where the training data and validation data still coincide), namely *have* + present participle and *have* + infinitive. For the other targets, the percentage of correctly predicted cases ranges from 4.29% with *get* + present participle to 87% with *make* + infinitive and passive *make* + *to*-infinitive. The tree shows broad tendencies, concerning 300 or more forms, but also small “niches”, which are applicable to just a few cases. Some of these should be discarded as irrelevant, but others can give us useful information about a particular structure. Thus, the rule “If PATIENT = no and CAUSEE = no, then Verb = MAKEINF (100%),” concerning only nine forms, points to particular constructions with *make* where neither the CAUSEE nor the PATIENT is expressed, cf.

- (1) Well I have in the past given them fifty each for Christmas but this year they’re going to have to **make** do with twenty. <BNC:S:KBF 4752>
- (2) This is direct experience, but it is not drama -- not until there is some pretence involved, some symbolic representation, some intention to **make** believe. <BNC:W:AM6 126>

As for the general tendencies, they tell us, for example, that if the CAUSER is inanimate and the CAUSEE is mentioned, *make* + infinitive is used in 78% of the cases, which represents 521 forms, cf.

- (3) And we also know that advertising reinforces smoking, it **makes** people think that smoking is okay. <BNC:S:FLM 346>

or that *have* and *get* + past participle are used in similar environments (viz. an animate CAUSER, a PATIENT, no CAUSEE and a volitional EFFECT), with the difference that *have* is chosen when the CAUSER is distinct from the (implied) CAUSEE (303 forms), while *get* is preferred when the CAUSER is co-referential with the CAUSEE or when the relation between the two participants is ambiguous (617 forms). Compare:

- (4) She is always going to the hairdresser and **having** her hair frizzed <BNC:S:KCN 4895>
- (5) But she couldn’t **get** the car started this morning. <BNC:S:KBY 50>

We also learn that some structures are used in more specific environments than other structures, which are more “versatile.” *Make* + infinitive is the structure that most often recurs as a leaf (6 leaves out of 14), which means that it can be used in many different contexts, includ-

⁶ This, incidentally, facilitates the decision tree building, as it would be an extremely slow and (for the computer) tedious process to predict the use of ten targets on the basis of some 40 variables.

⁷ In fact, the analysis of the individual variables reveals that the PATIENT, when expressed, presents almost identical characteristics with all four verbs (most notably, a predominantly inanimate nature, while the CAUSER and CAUSEE greatly differ in this respect).

ing the two contexts already alluded to, but also e.g. an animate CAUSER, a CAUSEE and a non-volitional EFFECT, as in:

- (6) Now you've **made** me forget what I was gonna say <BNC:S:KP2 256>

By contrast, the use of *make* + past participle is very restricted, as all the forms correctly predicted (i.e. 11 forms out of 16) are of the same type, viz. an animate CAUSER, a non-volitional EFFECT, a PATIENT and no CAUSEE, e.g.

- (7) The group who face the most difficulty getting adequately trained are women doctors, and they **made** their voices heard at the conference despite the organisers admitting with startling candour that they had not anticipated that this issue would arise. <BNC:W:CNA 215>

Now that we have a better idea of the kinds of results that are obtained by means of the decision tree, let us see how these results can be applied.

5. Applicability

5.1. Predictive power

Although the technique of decision trees is primarily meant to predict the membership of new cases, it should be noted that the immediate use of the predictive power of a tree is very restricted when it comes to linguistic phenomena. While it makes sense, say, to predict, for a new population, whether a person is a potential buyer of a particular product or not, or whether a credit card holder is likely to become an inactive customer (example given by Berry and Linoff, 1997: 268-273), it is of little use in itself to be able to predict the form that will be used in a given context. Except within the (necessarily artificial) frame of an exercise for learners, one would probably never be given a text from which all the causative verbs are missing and where the forms would have to be retrieved by means of an algorithm, for the simple reason that the verb is inseparable from the rest of the sentence. The (direct) usefulness of the predictive power of a decision tree in linguistics is therefore limited to the indication it gives of the reliability of the tree. However, decision trees become extremely powerful tools when they are applied in a more indirect way.

5.2. Scientific description of the use of causative constructions

What is useful in decision trees for our purposes is the set of rules that are automatically generated. Not only do they highlight the most relevant variables, but they also show which combinations of variables lead to the use of a particular verb. In other words, decision trees make it possible to describe the behaviour of causative constructions in real English and pinpoint the factors that influence the choice of one structure or another. As pointed out above, this is the sort of information that is most sorely lacking in the literature on causative constructions, although it is indispensable in order to make an informed choice, one that will sound natural and authentic.

This description, however, should be supplemented by other aspects, which are less adequately dealt with by means of a decision tree, but are nonetheless relevant to a description of the phenomenon under investigation. This is the case with the stylistic and lexical aspects of causative constructions. Introducing the speech/writing dichotomy in the decision tree is feasible, although it does not give rise to any significant improvement in the predictive power (74% of correctly predicted cases instead of 73%). As for genres, their great diversity makes the tree rather confusing, with rules such as "If genre is one of: conv disc hum med pol soc, then..." The whole stylistic aspect is therefore best described as an independent level, on a

par with the syntactic and semantic aspects described by the decision tree. The same applies to lexis. A collocational analysis of causative constructions reveals that each structure keeps strong preferential lexical company. Thus, the past participle with *make* frequently takes the form of *known* (18 occurrences in our data, that is 60% of all past participle constructions with *make*), *felt* (5 occurrences) or *understood* (2 occurrences), while with *get* and *have*, by far the most common EFFECT is *do* (almost 25% of all the EFFECTS with each causative). In addition, *get* frequently co-occurs with words referring to some sort of difficulty or effort, as in:

- (8) We're not picking on you we're just trying to **get** a conversation going here like, but it's very hard when you just sit there and say nothing. <BNC:S:KCX 6634>
- (9) Attempts to **get** parents to reduce calorific intake, if the cause of the obesity is psychogenic, are doomed to failure; possibly this is why these families are often so difficult to treat. <BNC:W:CGT 1461>

Again, introducing such elements would make the decision tree much more complex (also for the computer), resulting in rules of the type “If EFFECT is one of: *know feel understand*, then...” or “If context is one of: effort difficulty, then...” Therefore, it is perhaps preferable, for the sake of simplicity, to treat the lexical aspect of causative constructions separately from the decision tree.

5.3. Pedagogical grammar?

By alluding to simplicity in the preceding section, we are already moving towards another field of application for decision trees, namely pedagogy and more particularly second and foreign language acquisition. That English causative constructions are difficult to use for non-native speakers appears from various studies (e.g. Wong, 1983; Liu and Shaw, 2001 for Chinese learners; Altenberg and Granger 2001 for Swedish and French-speaking learners; or Helms-Park, 2001 for speakers of Hindi-Urdu and Vietnamese). And pedagogical grammars are of little help here, as they suffer from the same problems as the rest of the literature, namely incomprehensiveness and unreliability.

Although a decision tree built on the basis of authentic data can certainly provide learners with practical information to choose the most suitable causative structure in a given environment, it seems to us that the tree has to be adapted before it can be of real use to them. First, only the most relevant and basic variables should be taken into account, those which are at the learner's immediate disposal **before** constructing the causative structure (number of participants involved, animate or inanimate nature of the entities, degree of volitionality of the EFFECT). Second, it would be necessary to get rid of some of the uninteresting rules that account for only a few records (see above), as well as rules which make little sense (at least to students) such as “If CerEv is one of: no yes, then...” where the only possibility that is discarded is a “non-applicable” value for the nature of the CAUSER (event or not). Then, the causative structures that are not predicted by the automatic tree should somehow be taken into account in a “pedagogical decision tree.” This would involve allowing for more than one structure per leaf, with a note explaining the difference between the alternatives. The decision as to which additional structures to mention could be based on the proportion this structure represents, not vis-à-vis all the records classified by the rule in question, but vis-à-vis all the occurrences of this particular structure. To illustrate this, let us take the example of present participle constructions with *get*. The rule “If CAUSER = yes and CAUSEE = yes and PATIENT = no and CAUSER = animate and CAUSEE = inanimate, then...” provides a majority of MAKEINF records (116 out of 247, that is 46.96%). In an automatic decision tree, the leaf would therefore be labelled as MAKEINF. However, the records also include 64 instances of *get* + present

participle constructions. Although this is less than MAKEINF, these instances of GETPRP represent 49.61% of all the *get* + present participle constructions present in the corpus (64 constructions out of 129). So it is well worth mentioning in a pedagogical tree, since this rule shows one of the typical environments in which GETPRP occurs. More generally, one should systematise the tree as much as possible, so as to make it understandable to learners and easy to remember, even if it entails a decrease in precision. The resulting tree will thus have a weaker predictive power, but will be more effective in terms of second or foreign language acquisition.

6. Conclusion

Despite the complexity of causative constructions and the vagueness with which they tend to be described in the literature, it has been shown that the choice of a given causative verb and structure can be predicted to a certain extent by means of a decision tree. The main use of such a tree is that it emphasises the most relevant variables influencing this choice and, above all, the combinations of variables that lead to a particular choice. For such a tree to be usable in the field of second or foreign language acquisition, however, human intervention is needed in order to make the tree more economical and systematic.

References

- Altenberg B. and Granger S. (2001). The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics*, vol. (22): 173-194.
- Berry M.J.A. and Linoff G. (1997). *Data Mining Techniques for Marketing, Sales, and Customer Support*. Wiley and Sons.
- Comrie B. (1976). The syntax of causative constructions: cross-language similarities and divergences. In Shibatani M. (Ed.), *The Grammar of Causative Constructions*. Academic Press: 261-312.
- Dik S.C. (1980). The Dutch causative construction. In *Studies in Functional Grammar*. Academic Press: 53-89.
- Duhoux Y. and Lecoutre E. (2003). La prédiction de l'aspect verbal en grec ancien. In *Proceedings from Convegno Internazionale di Linguistica Greca*.
- Gries S.Th. (2003). Towards a corpus-based identification of prototypical instances of constructions. *Annual Review of Cognitive Linguistics*, vol. (1): 1-28.
- Guierre L. (1959). *The right Word ... Le Mot juste ... and the right Sound*. Librairie Vuibert.
- Helms-Park R. (2001). Evidence of lexical transfer in learner syntax. The acquisition of English causatives by speakers of Hindi-Urdu and Vietnamese. *Studies in Second Language Acquisition*, vol. (23): 71-102.
- Kemmer S. and Verhagen A. (1994). The grammar of causatives and the conceptual structure of events. *Cognitive Linguistics*, vol. (5): 115-156.
- Liu E.T.K. and Shaw P.M. (2001). Investigating learner vocabulary: a possible approach to looking at EFL/ESL learners' qualitative knowledge of the word. *International Review of Applied Linguistics in Language Teaching (IRAL)*, vol. (39): 171-194.
- McCawley J.D. (1968). Lexical insertion in a transformational grammar without deep structure. In Darden B.J., Bailey Ch.-J.N. and Davison A. (Eds), *Papers from the Fourth Regional Meeting of the Chicago Linguistic Society*, vol. (4): 71-80.
- Wierzbicka A. (1998). The semantics of English causative constructions in a universal-typological perspective. In Tomasello M. (Ed.), *The New Psychology of Language. Cognitive and Functional Approaches to Language Structure*. Lawrence Erlbaum Associates Publishers: 113-153.
- Wong S.C. (1983). Overproduction, underlexicalisation, and unidiomatic usage in the 'make' causatives of Chinese speakers. *Language Learning and Communication*, vol. (2): 151-163.