Linguistic Research 40(3): 353-385
DOI: 10.17250/khisli.40.3.202312.001

Cause identifiability and the causative alternation in English: A corpus-based analysis*

Hanjung Lee
(Sungkyunkwan University)

Lee, Hanjung. 2023. Cause identifiability and the causative alternation in English: A corpus-based analysis. *Linguistic Research* 40(3): 353-385. Many verbs in English show causative and noncausative uses. The goal of this paper is to verify whether the frequencies of the causative and noncausative uses of alternating verbs are modulated by the identifiability of the ultimate cause of the event in a given discourse context. On the basis of a corpus study of 12 change-of-state verbs, it will be shown that verbs that are predominantly used as a causative tend to have a cause argument with a higher degree of identifiability, while verbs that are more frequently used as a noncausative tend to have a cause argument with a lower degree of identifiability. Based on this evidence, I argue that cause identifiability in context is a crucial factor that more accurately characterizes verb-specific preferences for one variant over the other as well as the corpus frequency distribution of the alternation variants than other lexical semantic factors discussed in the literature. **(Sungkyunkwan University)**

Keywords causative alternation, causative variant, cause identifiability, communicative efficiency, noncausative variant

1. Introduction

Many verbs in English show causative and noncausative uses. This possibility−known as the causative alternation−is illustrated in (1) with the verb *break*. In English verbs that participate in this alternation show both uses such that the causative use has roughly the meaning ‘cause to V-intransitive’. Concomitantly, the causative variant includes a causer

* This research was supported by a grant from the Ministry of Education and the National Research Foundation of Korea (NRF2023S1A5A2A01074277). I would like to thank two anonymous reviewers for helpful comments and Jiyoun Kim for assistance with data collection. I am particularly grateful to Prof. Minhaeng Lee for introducing useful tools for corpus analysis used in this paper and providing help with many aspects of this research. It goes without saying that all inadequacies and errors are entirely my own.

argument that the noncausative variant lacks.

(1) a. Mary broke the vase. (causative variant)
b. The vase broke. (noncausative variant)

In the English example above, the change between the causative and noncausative variant does not result in a formal change in the verb.[1] Cross-linguistically, however, the causative alternation is encoded in ways that involve a formal change in the alternating verb, though the same semantic relationship between the variants is maintained.

In past decades, the causative alternation has been the focus of many studies in various subfields of linguistics, such as linguistic typology (Nedyalkov and Silnitsky 1973; Haspelmath 1993, 2016; Nichols, Peterson and Barnes 2004; Comrie 2006; Haspelmath et al. 2014), theoretical linguistics (Levin and Rappaport Hovav 1995; Reinhart 2002, 2016; Alexiadou et al. 2006; Koontz-Garboden 2007, 2009; Schäfer 2008; Rappaport Hovav and Levin 2012; Rappaport Hovav 2014), and corpus linguistics (McKoon and Macfarland 2000; Wright 2001, 2002; Heidinger 2015; Samardžić and Merlo 2018). The causative alternation is an especially attractive topic of research because it allows insights into complex constraints underlying the linguistic coding of causal events within and across languages and the consequences of such constraints for language structure and use.

Most theoretical accounts of the causative alternation to date have focused on (i) lexical properties of alternating verbs, (ii) systematic relation between causative and noncausative uses of the alternating verbs, and (iii) links between alternating verbs and the alternation variants (see Levin (2015) for a review). More recent accounts of this alternation recognize complexity of the issues (i)-(iii), taking into account both lexical semantic properties of verbs and the contextual conditions that determine which variant is chosen when, in principle, both might be possible. For example, Rappaport Hovav and Levin (2012) and Rappaport Hovav (2014) show that in addition to semantic properties of verbs, contextual factors must also be key components of any account of the alternation. Based on previously unobserved contextual differences between causative and noncausative uses of verbs, Rappaport Hovav (2014) demonstrates convincingly that

1 The two variants or alternants of the causative alternation are referred to in various ways. Following Levin (2015), I refer to them as causative and noncausative variants to avoid labels such as (in)transitives and anticausatives that reflect on the morphosyntactic form of the verb in this variant.

contextual factors such as the identifiability of a cause argument, the discourse context, and speakers' intentions and perspectives on the situation being described are at play in the relation between the causative alternation variants, and these factors determine the variant that is appropriate as a description of a given situation.

The crucial role played by contextual factors in determining the choice of a variant in a given context raises a few questions that remain unanswered in the extensive body of research on the causative alternation. Alternating verbs may differ with respect to how often they are used as a causative and as a noncausative. For example, *open* is used more often as a causative, while *melt* is found more often as a noncausative (Haspelmath et al. 2014). Given verb-specific preferences for one variant over the other, questions arise as to (i) why verbs differ in their frequency of use in the two variants and (ii) what contextual factors explain verb-specific preferences for one variant over the other. The frequencies of verb uses have been studied in relation to morpho-syntactic encoding of alternating verbs in crosslinguistic studies (Haspelmath 1993; Haspelmath et al. 2014; Heidinger 2015; Samardžić and Merlo 2018). Nevertheless, the systematic consideration of contextual factors is a very recent development in the study of the causative alternation as well as the linguistic coding of causal events within and across languages, and so far very little empirical work has been conducted in this domain. The goal of this paper is to verify whether the frequencies of the causative and noncausative uses of alternating verbs are modulated by the identifiability of the ultimate cause of the event in a given discourse context.

This paper is structured as follows. Section 2 provides a review of theoretical approaches to the causative alternation in English, documenting a shift from accounts that are semantically-oriented in nature to accounts that acknowledge pragmatic components to their analysis. Section 3 first discusses the notion of cause identifiability, and then formulates specific predictions tested in the analysis. Section 4 is the empirical core of this paper and devoted to a corpus study designed to test the predictions against a total of 2400 instances of causative and noncausative uses of 12 alternating verbs extracted from the automatically parsed British National Corpus (BNC). Careful scrutiny of the corpus data will show that the causative and noncausative uses of the verbs differ in the identifiability of the ultimate cause of the event, i.e., the probability of the event described by the verb and its argument having a clear identifiable cause. Based on this evidence, in section 5, I argue that cause identifiability is a crucial factor that accounts for variation in the English causative alternation and that other lexical semantic factors

discussed in the literature such as spontaneity or the involvement of an external causer are less successful in explaining the corpus frequency distribution of the alternation variants and verb-specific preferences for one variant over the other. Section 6 concludes the paper.

2. Theoretical approaches to the causative alternation

Theoretical approaches to the causative alternation fall into two major categories. The dominant approach has been a semantic one which attempts to account for the alternation in terms of lexical-semantic properties of verbs. The second approach attributes the use of causative and noncausative variants to contextual factors, stressing the interplay of lexical, semantic and contextual factors in characterizing the relation between the two alternation variants. This section examines these two classes of approaches.

Lexical-semantic approaches to the causative alternation have attempted to capture the relation between verbs and the alternation variants by positing lexical distinction between subclasses of verbs and mapping different verb classes onto distinct argument realization patterns which are assumed to correspond to distinct types of eventualities. This class of accounts has been most fully developed in Levin and Rappaport Hovav (1995) and Reinhart (2002, 2016).

Levin and Rappaport Hovav (1995) propose that the crucial lexical-semantic distinction for characterizing the class of alternating verbs is between verbs denoting internally caused eventualities and those denoting externally caused eventualities. For example, verbs such as *bloom* and *blossom* can be considered to describe internally caused eventualities, because natural events like blooming and blossoming arise from internal properties of the blooming or blossoming entity; these events do not require a causer external to the blooming or blossoming entity. On the other hand, verbs such as *break* and *open* can be considered to describe eventualities that are externally caused, i.e., caused by an entity external to the broken or opened entity. Levin and Rappaport Hovav (1995) assume that verbs denoting internally caused eventualities are lexically monadic, selecting only the argument denoting the entity undergoing the change specified by the verb, and verbs denoting externally caused eventualities are lexically dyadic, selecting in addition a cause argument. A subset of verbs denoting externally caused events undergo a process of lexical binding of the external argument, preventing the external argument

from being expressed syntactically, thereby resulting in the noncausative variant of the alternation.

Reinhart (2002, 2016) provides a similar characterization of the class of alternating verbs. Like Levin and Rappaport Hovav's (1995) analysis, her analysis of the causative alternation takes the causative variant to be basic, deriving the noncausative variant from it via a lexical operation of decausativization. This operation applies to verbs whose thematic structures involve two arguments, which have the lexical specification [+c(ause change)] and [–c, –m(ental state)], respectively:[2]

(2) V ([+c] (=underspecified cause), [−c, −m] (=theme))

The [+c] specification identifies arguments that fall under a broad notion of cause, subsuming agents, natural forces, and instruments (Reinhart 2016: 25). The decausativization operation eliminates the [+c] argument altogether, giving a thematic structure with a single argument, the theme:

(3) Decausativization (Reinhart 2016: 27, (52))
V_{Acc} ($\theta_{[+c]}$, θ_j) → V (θ_j)

The [+c] specification on Reinhart's account singles out verbs that simply specify that their subject is a cause. It is thus parallel in many ways to the absence of a lexical specification for the causing event on Levin and Rappaport Hovav's (1995) account, external causation with no specification of the causing subevent: both accounts characterize the class of alternating verbs over the causative variant of the alternation, predicting that alternating verbs can appear with a wide range of semantic types of NPs, as shown in (4a) and (4b). In contrast, the verb *murder* specifies something about the causing event (it must involve intention). Hence, this verb cannot undergo lexical binding or decausativization, and does not participate in the alternation, as shown in (5).

2 Reinhart's (2002, 2016) theta system assumes two binary features, *cause change* and *mental state*. The possible combinations of these two features define four clusters: [+c, +m], [–c, +m], [+c, –m], and [–c, –m]. The role that each of these clusters corresponds to is shown below.

Lexical specification	[+c, +m]	[–c, +m]	[+c, –m]	[–c, –m]
Role	agent	experiencer	instrument cause	theme patient

(4) a. The vandals/the rocks/the storm broke the windows.
b. The butler/the key/the wind opened the door.
(5) a. The hit men/*the bullets/*the plan murdered the gangster.
b. *The gangster murdered.

However, a range of subsequent work, particularly, Rappaport Hovav and Levin (2012) and Rappaport Hovav (2014) shows that the characterization of the class of alternating verbs provided by Levin and Rappaport Hovav (1995) and Reinhart (2002, 2016) is not accurate, and meshes better with the position that alternating verbs are lexically associated only with their internal argument(s) (Rappaport Hovav 2014: 13).

What is of particular interest here is the second type of motivation for taking the causative variant to be basic – the purported asymmetry in the availability of the causative alternation – and the nature of factors which constrain the alternation. In what follows, I take issue with the common claim made by Levin and Rappaport Hovav (1995) and Reinhart (2002, 2016) that the conditions which restrict the causative alternation are lexical, a claim that Reinhart (2002, 2016) develops more explicitly than Levin and Rappaport Hovav's earlier analysis.

If it is true that the lexical specification of the verb is all that constrains the alternation, then all externally caused verbs which are compatible with NPs bearing a range of semantic roles in subject position are expected to appear in both variants irrespective of the properties of the NP chosen as theme. This is, however, not the case. Change of state verbs that may select agents, natural forces, and instruments as subjects may nevertheless not have noncausative uses for some choices of causative variant theme. As Levin and Rappaport Hovav (1995) point out, causative sentences of change-of-state verbs−the prototypical causative alternation verbs−lack a noncausative counterpart, as in (6) and (7).

(6) a. He broke his promise/the contract/the world record.
b. *His promise/*The contract/*The world record broke.
(Levin and Rappaport Hovav 1995: 85, (9))
(7) a. The waiter cleared the table.
b. *The table cleared. (Levin and Rappaport Hovav 1995: 104, (55))

The asymmetry in available variants illustrated in (6) and (7) is pervasive. As

Rappaport Hovav and Levin (2012) show, comparable patterns are found with other change of state verbs such as *empty, lengthen* and *shorten*. These verbs meet the lexical specification of the class of alternating verbs, and they in fact do alternate, as shown in (8)-(10). However, there are certain cases where a noncausative variant is ill-formed in isolation, as in (7b), (11b) and (12b).

(8) a. The wind cleared the sky.
b. The sky cleared. (Levin and Rappaport Hovav 1995: 104, (56))
(9) a. The mad scientist lengthened the days.
b. The days lengthened. (Levin and Rappaport Hovav 1995: 105, (58))
(10) a. I emptied the tub.
b. The tub emptied. (Rappaport Hovav 2014: 12, (20))
(11) a. The dressmaker lengthened the skirt.
b. *The skirt lengthened. (Levin and Rappaport Hovav 1995: 105, (57))
(12) a. I emptied the trash can.
b. *The trash can emptied. (Rappaport Hovav and Levin 2012: 158, (13))

Levin and Rappaport Hovav (1995: 102) suggest that the rule of lexical binding is restricted to cases of verbs that denote an "eventuality [that] can come about spontaneously without the intervention of an agent" (cf. Smith 1978; Haspelmath 1993). This accounts for the contrast between (8b)-(10b) which describe events that can take place without an agent on the one hand, and (6b)-(7b) and (11b)-(12b) which cannot, on the other.

However, the spontaneity condition on application of lexical binding is problematic on both theoretical and empirical grounds. First, notice that it is not the meaning of the verb which determines spontaneity or involvement of an agent. Rather, it is world knowledge and properties of eventualities which tell us that certain events (e.g., clearing of the sky and lengthening of the days) do not need the intervention of an agent, while others (e.g., clearing of the counter/the table and lengthening of the skirt) can only come about with the intervention of an agent. The constraint on the availability of variants then does not seem to be lexical in nature (Schäfer 2009; Rappaport Hovav and Levin 2012). Rather, as Rappaport Hovav (2014) correctly points out, it appears to be a constraint on the appropriateness of the noncausative variant in the description of events of a certain type. The nonlexical nature of the constraint on the alternation, then, undermines the

basic assumptions of Levin and Rappaport Hovav's (1995) and Reinhart's (2002, 2016) accounts that all alternating verbs are inherently dyadic and the noncausative form is derived through a lexical rule which removes the cause argument.

Another serious problem of the spontaneity condition on application of lexical binding is the preponderance of noncausative uses of alternating verbs which describe events involving an external agent that is not contextually identifiable. Here are three illustrative examples of noncausative sentences of *open* from the BNC. These examples describe events of opening which require the participation of an agent: opening shops, a chapter, and an inquest; these contrast with opening a door which could be done by an agent, an instrument, or even a natural force (see (4b)).

(13) a. It is hoped that the builders will be on site by July and the first shops will open by Christmas 1994. (BNC miscellaneous texts)
b. The chapter opens with an introduction to the semiclassical theory of light-atom interactions and optical resonators. (BNC academic prose: natural sciences)
c. An inquest will open today. (BNC news script (TV autocue data))

In (13), the addressee is not able to infer the identity of the agent, but rather can infer the type of agent: the owner of a shop, the author of the chapter, and the coroner. In such examples as these, the unexpressed agent is inferrable, but has a referent whose exact identity is unidentifiable and unimportant. This kind of inferrable agent is dispreferred in causative uses, if not excluded. However, contrary to the fact, Levin and Rappaport's (1995) account predicts that agentive uses of verbs lack a noncausative variant, as on their account a process of lexical binding is restricted to cases of verbs denoting an eventuality that can come about spontaneously.

Rappaport Hovav and Levin (2012) and Rappaport Hovav (2014) propose a non-derivational approach to the alternation which diverge from their own earlier analysis and that of Reinhart (2002, 2016), as well as non-derivational analyses embodied in Alexiadou, Anagnostopoulou and Schäfer (2006) and others. They provide systematic evidence in support of the view that in English all alternating verbs are lexically associated with their patient only. The cause, then, is introduced non-lexically. A central constraint on the addition of the cause argument is the well-known constraint that the cause must be construable as a direct cause – a cause with immediate control over the

eventuality. Building on Rappaport Hovav and Levin (2012), Rappaport Hovav (2014) suggests that a causer can be realized with a change-of-state verb just when the event being described involves direct causation in the sense defined by Wolff (2003: 5). This constraint contributes to the determination of the semantic type of NPs allowed as the external argument in given contexts.

Rappaport Hovav (2014) then explores the contextual factors which license or even require the noncausative variant, confronting the question of which variant is appropriate in a given discourse context. Drawing on Grice's maxims of conversations, she argues that if the expression of the cause is deemed relevant, it will be preferred in most cases because the causative variant is more informative. The noncausative is licensed or even required either if the cause is recoverable in some way or if the speaker does not know what the cause is. Rappaport Hovav (2014) shows that this account covers constraints on the alternation which have been taken to be lexical and also those which are more clearly nonlexical in earlier accounts.

To summarize, this section has shown that the English causative alternation is not constrained only by the meaning of verbs but rather the distribution of the variants is constrained by both lexical and contextual factors. Recent accounts of this alternation recognize this complexity, explaining both the relation between verbs and the variants and the contextual conditions that determine which variant is chosen when, in principle, both might be possible. Rappaport Hovav's (2014) study is an important step in this direction which opens up new perspectives on the causative alternation.

To the best of my knowledge, there are no studies that have tested the empirical validity of this new approach to the causative alternation on the basis of a systematic analysis of naturally occurring data. To fill this gap in the literature, this paper attempts to verify against corpus data whether the frequencies of the causative and noncausative uses of alternating verbs are modulated by the identifiability of the ultimate cause of the event in a given discourse context.

3. Cause identifiability and its predictions

This section first discusses the notion of cause identifiability, and then formulates specific predictions tested in the analysis.

For the purpose of the empirical analysis, cause identifiability is defined as a gradable

property of an event that depends on the extent to which the event expressed by the verb and its argument has a clear identifiable cause. Figure 1 summarizes the classification of cause types according to the ease of identification.

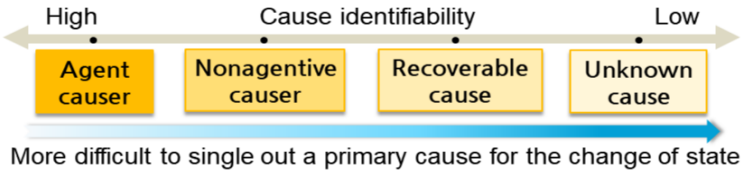

Figure 1. Classification of cause types according to the ease of identification

Here, the scale of cause identifiability can be conceived of as being derived from the degree to which the primary cause for a change of state can be singled out. Cause types higher on this scale are more easily identifiable while it is more difficult to uniquely identify cause types lower on the scale.

AGENT CAUSER. Agents are considered the prototypical cause (Talmy 1976, 2000; Lakoff 1990; Croft 1991). They are typically easily isolatable as an ultimate cause because they often have intentions to bring about specific changes. When an agent is involved in causing an event, that agent, as the ultimate cause of the event, must be usually be expressed because "it is more informative to mention the ultimate cause of a change of state, if the ultimate cause is known" (Rappaport Hovav 2014: 26). This appears to be why the noncausative variant is generally dispreferred as a description of an agentive situation unless the agent was established previously in the discourse or is inferrable from the context.

NONAGENTIVE CAUSER. When an event is nonagentive, however, it may be more difficult to identify its ultimate cause, and noncausative variant is felicitous. Rappaport Hovav (2014) argues that sentences such as *The window opened* and *The vase broke* are easier to accept compared to sentences such as *The table cleared* and *The skirt lengthened* because it is easy to imagine a scenario in which the speaker sees the change of state but not the act which brings it about: i) the change of state is brought about by some nonagentive cause which lacks intentions (e.g., by the wind), and then the cause may not be easily identifiable; (ii) the change of state is brought about by the action of an agent from a distance.

RECOVERABLE CAUSE. When there are a number of causes normally present

simultaneously, it becomes even more difficult to identify the ultimate cause of an event. This happens in various circumstances when the cause is recoverable by default. According to Rappaport Hovav (2014: 23), many natural events of change have default causes. As mentioned in section 2, whether a verb describe such a change of state is not always lexically determined by the verb, but often by the verb and its argument(s) and sometimes only in particular contexts. Consider, for example, a verb such as *lengthen*. Hair lengthens in the normal course of events, and so do days, but skirts do not. The cause of the lengthening of the days or hair is normally not mentioned because though not all speakers have scientific knowledge of what causes days or hair to lengthen, it is known that there is a given set of causes which does not normally change. This is the sense in which the cause is recoverable by default. As Rappaport Hovav (2014) points out, when the cause is recoverable by default, the use of the causative would be odd, and this can be attributed to some version of the Maxim of Manner which dictates avoiding prolixity.

A cause can also be recoverable when it has been established in the preceding context. These are cases such as those illustrated in (14) below, where the cause of the change mentioned in the preceding context licenses the noncausative variant of the verb.[3] In (14a) the causing action was carried out unintentionally, while in (14b) it was carried out intentionally.

(14) a. I leaned against the door and it accidentally opened.
(Rappaport Hovav 2014: 17, adapted from example (41))
b. I pushed and pushed on the door, and it finally opened.
(McCawley 1978, cited in Rappaport Hovav 2014: 22, (62))

3 In this regard, the discourse function of noncausatives can be seen to overlap partially with that of "short" passives (passives without a *by*-phrase). A corpus study of Thompson (1987) has shown that the short passive is used when the agent is not to be mentioned explicitly because its identity is unknown, unimportant, or evident from the context. Yet, the noncausative and the short passive are essentially different constructions and need to be analyzed differently. As discussed by Wanner (1999), the function of the passive cannot be to maximally background the external argument, but to background it to a certain extent to an element with specific syntactic and semantic characteristics, what she calls the implicit argument. The effect of the implicit argument in passives is illustrated in (i). Unlike noncausatives, (verbal) passives are compatible with agent-oriented adverbs like *(un)willingly, deliberately*, and *voluntarily*, even if there is no explicit agent in the sentence:

(i) a. The price was decreased unwillingly.
b. *The price decreased unwillingly.

A different case of recoverability is when the cause may be hinted at in the surrounding context. An example would be (15) uttered in the context of an event with many tables and food served by waiters. In such instances, a noncausative variant may be used to describe the event.

(15) As the night wore on, the tables slowly cleared and there was nothing left for the late comers to eat. (Rappaport Hovav 2014: 26, (81))

UNKNOWN CAUSE. The noncausative variant may also be appropriate when the speaker does not know the ultimate cause of the event. As McCawley (1978) points out, the choice of the verbs *come* and *go* determines the orientation of the speaker: in (16) the speaker is most likely in the lunchroom. As a consequence, this example invites the inference by the hearer that the speaker does not know how the door opened.

(16) The door of Henry's lunchroom opened and two men came in. (McCawley 1978, cited in Rappaport Hovav 2014: 22, (59a))

A different case of unknown or unidentifiable cause is the cause for a gradual change discussed by Rappaport Hovav (2014). Consider (17):

(17) With the 1929 stock market crash, skirts lengthened but kept their narrow silhouette, with longer waistlines. (Rappaport Hovav 2014: 27, (83))

This sentence does not report a change in the length of an individual skirt, but rather in the kind skirt. It reports a situation in which over the course of time, there are a variety of instantiations of the kind and when comparing them one to another over the course of time the length increases. There are probably a variety of causes (and reasons) for the gradual change in the kind, but they are likely to be diffuse and not uniquely identifiable.

PREDICTIONS. In this paper, taking the proposed scale in Figure 1 above as a basis for my analysis of the causative alternation, I attempt to empirically test it against corpus data. My central hypothesis is that the ease of cause identification correlates with frequencies of verb uses. This hypothesis predicts a covariation between the ease of cause identification and the frequencies of verb uses. The specific predictions that will be tested

in this study can be formulated as follows:

(18) Predictions
  a. A positive correlation exists between the ease of cause identification and the frequency of causative uses.
  b. A negative correlation exists between the ease of cause identification and the frequency of noncausative uses.

The basic expectation is that the two variables—cause identifiability and verb uses—covary in two different ways: first, for causative uses, a positive correlation between the ease of cause identification and the frequency of causative uses is predicted; secondly, for noncausative uses, a negative correlation between the ease of cause identification and the frequency of noncausative uses is predicted. According to this prediction, a high degree of cause identifiability increases the likelihood that the causative is used, and a low degree of cause identifiability increases the likelihood that the noncausative is used. The following section discusses a corpus study which tested these predictions against corpus data.

4. Empirical study

This section first discusses procedures of data collection and analysis (section 4.1) and then presents the results of the data analysis (section 4.2).

4.1 Procedures

4.4.1 Selection of verbs

To test the prediction that there is a correlation between the ease of cause identification and the frequencies of uses of alternating verbs, the following 12 verbs were analyzed.

Table 1. Sets of alternating verbs

Set A	Set B
dissolve, dry, fill, freeze, melt, sink	*break, clear, close, empty, lengthen, open*

An important consideration in the selection of the verbs was to have sets of verbs which differ in frequencies of verb uses. However, since relative frequencies of causative and noncausative uses of alternating verbs have not yet been systematically investigated for English, frequencies of verb uses could not be used for verb selection. An alternative criterion used for verb selection was relative frequencies of the morphological marking of the alternation. An influential study by Haspelmath (1993) provides a systematic typological analysis of the morphological marking in the two variants of alternating verbs across a wide range of languages. In Haspelmath's (1993) view, the morphological marking of the alternation reflects the salience of a cause. In a series of work (Haspelmath 1993; Haspelmath et al. 2014; Haspelmath 2016), he provides quantitative evidence for this claim by showing that the relation between the morphological marking and the salience of a cause cannot be observed in any single language (or a small set of languages), but it becomes apparent when a large sample of languages is observed. For example, verbs that are the lexical equivalents of English *dry, freeze*, or *melt* tend to be marked when used causatively in many different languages (causative marking), whereas the equivalents of English *open, break*, or *close* tend to be marked in their noncausative uses (anticausative or noncausative marking).

Thus, under the assumption that relative frequencies of the morphological marking of the alternation is an indicator of the salience of a cause, the two sets of verbs were compilated: Set A and Set B. Set A includes six verbs with a low frequency ratio between noncausative (N) and causative (C) marking, i.e., the N/C ratio values lower than 2 in Haspelmath's (1993) study.[4] Set B includes three verbs with the higher N/C ratio values (greater than 8) in Haspelmath's (1993) study: *break, close* and *open.* Set B include three other verbs that are not included in Haspelmath's (1993) study: *clear, empty* and *lengthen.* These are deadjectival change-of-state verbs which have causative uses with

4 The N/C ratio in Haspelmath (1993), i.e., the ratio of noncausative marking to causative, is calculated as shown below:

$$\text{N/C(verb)} = \frac{frequency\ of\ noncausative\ marking}{frequency\ of\ causative\ marking}$$

The verbs with a low N/C ratio describe events that are likely to happen without the involvement of an agent or any external force. According to Haspelmath (1993), if these verbs are used with an overt agent, their form contains a morphological marker in the majority of languages (causative marking). The verbs with a high N/C ratio typically specify an agent: If the agent is not specified, the verb tends to get some type of morphological marking across languages (anticausative or noncausative marking).

agentive subjects such as *The waiter cleared the table*. These verbs were selected because it is desirable to have variation in the class of cause-salient verbs. As mentioned in section 2, causative uses such as *The waiter cleared the table* lack noncausative counterparts, i.e., **The table cleared*. Such noncausatives with deadjectival verbs are judged ill-formed in isolation and need more contextual support than noncausatives with verbs like *break* and *open* (e.g., *The vase broke*) to license the noncausative in an agentive use (Rappaport Hovav 2014: 26).

Although the sets are too small to systematically detect the impact of cause identifiability on a wide range of verb classes, the limitation to 12 verbs has practical reasons, namely the fact that the type of cause for a change of state described in all relevant data (a total of 2400 examples) had to be determined on the basis of careful examination of uses of verbs in context. The following subsection discusses how causative and noncausative uses of the 12 verbs have been identified and extracted.

4.1.2 Extraction and coding of examples of verb uses

Causative and noncausitive instances of the verbs were extracted from the BNC, a 100 million word collection of samples of written and spoken language from a wide range of sources originally created by Oxford University Press in the 1980s – early 1990s. The BNC was chosen as the data source for a first step toward multi-corpora examination, which will compare the use of the same verbs with that in other corpora of English representing more recent language use (e.g., the BNC2014 and the ANC).

Syntactic relations needed for determining whether a verb is realized as causative (with a direct object) or as noncausative (without a direct object) are identified on the basis of automatic parsing using Stanza, an open-source Python natural language processing toolkit supporting 66 human languages developed by the Stanford NLP Group (Qi et al. 2020). In this study, only instances of active transitive realizations and intransitive realizations will be considered. Although they are basically transitive realizations, the passive instances were excluded because the difference between active and passive transitive uses and between passives and intransitive uses is crucial for the causative alternation, as discussed in detail by Wanner (2009). This result is obtained by using a corpus concordance system called the Corpus Query Processor (CQP) supported in Corpus Workbench (CWB). CWB is a widely-used architecture for corpus analysis,

originally designed at the Institut für Maschinelle Sprachverarbeitung (IMS), the University of Stuttgart (Christ 1994). As explained by Evert and Hardie (2011), Hardie (2012) and Lee (2015), CQP has a number of advantages as a tool for corpus analysis, including its implementation of a flexible search language that allows complex query patterns and web-based corpus interfaces. It also provides a range of useful query postprocesses.[5]

Accordingly, this study collected the data from the BNC text comprising sentences automatically parsed with a dependency parser by employing the CQP query language, i.e., regular expression patterns which make reference to conditions for the search (e.g., part-of-speech, word forms, dependency relations, token sequences, etc.).[6] For all verbs, the corpus queries led to far too many hits that fit the specifications for the search. For practical reasons, only a total of 2400 instances (200 instances per verb) were considered and manually coded for cause type. This number does not include the irrelevant examples that had to be sorted out.

Before moving on to discuss how the notion of cause identifiability was operationalized in the actual coding procedure of the corpus data, it is worth outlining some of the decisions which need to be made in order to single out irrelevant data.

The first decision concerns treatment of processing-related errors. Since all the data used in this study are collected automatically from an automatically parsed corpus, they necessarily include such errors. I have performed a manual evaluation of the pre-final data set to assess the degree to which they correspond to the correct analyses. The validation shows that there is a bias toward labeling the extracted instances as noncausatives when they are not and that such a bias is found in instances of all verbs included in the analysis. The noncausative bias was mainly due to the fact that English passive forms headed by the verb *be* such as *The door was opened* were not recognized

5 For a comprehensive introduction to the query language used within CQP, its use of regular expressions and more sophisticated query options, see Hoffmann et al. (2008: 217-245), Lee (2015: 61-104) and the CQP Query Language Tutorial available from http://cwb.sourceforge.net/documentation.php. See also Lee (2015) for a useful introduction to the use of the CQP query language for searching the parsed BNC for complex lexico-grammatical patterns. He further presents quantitative analyses for the collected data, including collostructural analyses of certain subtypes of causative constructions.

6 Dependency parsing builds a tree structure of words from the input sentence, which represents the syntactic dependency relations between words. The resulting tree representations, which follow the Universal Dependencies formalism (https://universaldependencies.org), are useful in many downstream applications. In Stanza, dependency parsing is performed by the DepparseProcessor (Qi et al. 2020). For detailed discussion of the Universal Dependencies formalism and dependency parsing, see de Marneffe and Manning (2008/2016), Manning et al. (2014) and Lee (2021).

as passives, but they were identified as noncausatives instead. Other cases of incorrect analyses were due to parsing errors which involve assignment of an incorrect form. For example, the intransitive form was assigned to the instance of the transitive form with an elided object, or the actual forms found in the instances were not verbs but were adjectives or nouns (e.g., *open, close, clear, dry*). As the current extraction method cannot deal with these cases automatically at the moment, such errors were manually removed from the final data set.

Further, I have limited myself to one-word uses of the verbs, and excluded complex verbs made up of [verb + particle], such as *sink in, dry up*, and *melt away*. The reason for this was that some multi-word uses of verbs under scrutiny do not show the causative alternation. For example, the verb *break* is frequently used in the verb particle construction (VPC), and show some concentrated use as a sense that does not have a change-of-state interpretation, as in *the war broke out*. The VPC *break out* has a coming-into-existence sense and lacks the causative counterpart with a sense paraphrasable as 'cause the war to happen'. As it is difficult to exclude only such VPCs, all VPCs were excluded from the analysis.

A final remark concerns the polysemy of the verbs. All verbs under investigation here are attested in the corpus data both in concrete and figurative uses (e.g., *the lake froze* (concrete use) vs. *freeze assets* (metaphorical use)). As a basic principle, both uses were considered, so long as they were extensions of the concrete meaning, rather than uses which no longer preserved any aspect of the physical sense.

After the irrelevant data had been excluded, the relevant data were manually coded for cause type to form the final data set for this study. The cause types coded in the data are summarized in Table 2.

Table 2. Summary of cause types in different degrees of identifiability

Cause type		Identifiability
Identifiable cause (IC)	Agent causer (AC) (exs. (6a), (7a))	High ↑
	Nonagentive causer (NC) (ex. (8a))	
Recoverable cause (RC)	Previously mentioned cause (ex. (14))	
	Hinted cause (ex. (15)); default cause (exs. (8b), (9b))	
Unknown cause (UC)	a cause from a distance (ex. (16)); a cause for a gradual change (ex. (17)); an unidentifiable cause (exs. (13), (19))	↓ Low

One complication in the coding of the data comes from the fact that the cause types in Table 2 co-occur in real language use. It is possible that a cause is nonagentive and recoverable, as illustrated in (14a) above, and agentive and recoverable, as illustrated in (14b). An example from the BNC in (19) below illustrates the same point. The *when*-clause in this example describes closing of an hospital brought about with the intervention of an agent that is not clearly identifiable in the context. We can thus say that the cause for this change of state is an instance of both the agent causer and the unknown cause.

(19) Careful plans have been made for these people so that when the hospital eventually closes they will not find themselves on the streets.
(BNC non-academic: politics, law, education)

Despite the possibility of co-occurrence in real language use, the four cause types (AC, NC, RC and UC) were coded here in ways that they do not overlap. This decision was motivated by the need to examine whether and how the frequencies of the uses of the investigated verbs differ in the degree of identifiability of the ultimate cause of the event in a given discourse context.

For this purpose, an identifiable cause (IC) is defined as a cause of a change of state that is explicitly singled out as the primary cause. One such way is by means of a lexical causative. The literature on lexical causatives has determined that the subject of the causative is conceptualized as proximate and ultimate (Wolff 2003; Levshina 2022).[7]

The coding procedure of the data reflected this widely accepted view by counting causes realized as the subject in the causative variant as instances of identifiable cause. Identifiable causes construed as capable of causing a change with volition have been

7 A lexical causative is a transitive verb signifying causation. It has often been noticed that lexical causatives apply to cases of direct causation but not to extended chains of causations. Indirect causal chains can, by contrast, be expressed in a periphrastic causative, in which a verb is embedded as a complement of make or some other causal verbs like *have, cause* or *let* (Fodor 1970; Shibatani 1976; Pinker 1989; Wolff 2003). As observed by Dixon (2000), more or less compact causatives vary not only with regard to (in)directness of causation, but also with regard to other parameters, such as involvement of the causer in the caused event and the causer's intentions. Levshina (2022) presents evidence based on typological data, corpora, and an artificial language learning experiment which demonstrates that the features of causative situations expressed by less compact forms across languages are less frequent than the features represented by more compact forms. Developing the line of argumentation in Haspelmath (2008, 2016), she argues that this pattern arises due to the principle of efficient language use, which involves negative correlation between accessibility and costs.

coded as agent causers (ACs) whereas ones that lack volition have been coded as nonagentive causers (NCs).

The restriction to the subject of lexical causatives excludes causes specified or hinted in the context from cases of identifiable cause. Instead, such causes were coded as recoverable causes (RCs). Alternatively, causes can also be recoverable because they are part of what discourse participants know about the way the world works: they are recoverable by default. Examples of a default cause of the event are causes of changes which happen in the normal course of events such as lengthening of hair, growing of a tree and clearing of the sky. According to Rappaport Hovav (2014: 24), one property which distinguishes default causes from identifiable causes is that there are a number of causes normally present simultaneously for changes which come about in the normal course of events. For this reason, it is typically difficult to identify a single, isolatable cause for such changes.

Causes can also be difficult to identify when a change occurs gradually over the course of time illustrated in (17) above. As Rappaport Hovav (2014: 27) point out, there are typically a variety of causes for the gradual change in the kind, but they are not uniquely identifiable. Such cases have been coded as instances of an unknown cause (UC), along with cases such as the example in (19), where the cause is not identifiable due to insufficient information in context.

4.2 Results

This section presents the results of the corpus study of the BNC data. Beginning with verb uses, Figure 2 shows the percentage of the two variants for each of the 12 verbs.

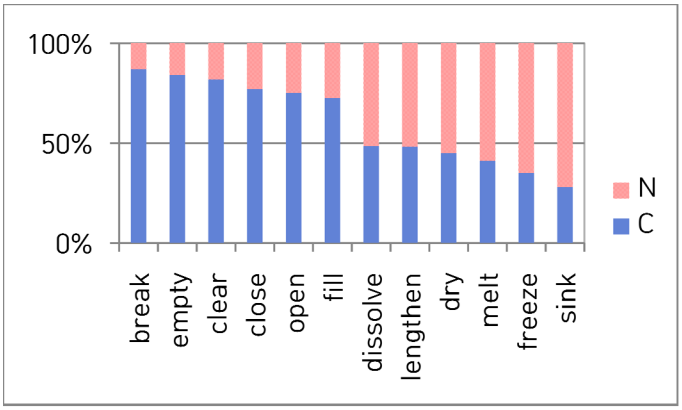

Figure 2. Proportion of verb uses in the two variants

As Figure 2 shows, there is a gradation in the two variants or alternants. 6 verbs in the data, i.e., *break, empty, clear, close, open* and *fill,* are predominantly used as a causative whereas 4 verbs, i.e., *dry, melt, freeze* and *sink,* are more frequently used as a noncausative; *dissolve* and *lengthen* occur with similar frequency in the two variants. Crucially, the first class of verbs has a higher percentage of agent causer (AC), while the latter class has a higher percentage of recoverable cause (RC) and unknown cause (UC). As shown in Figure 3, the proportion of agent causer decreases steadily from left to right, while the proportion of recoverable cause and unknown cause increases from left to right. A further notable result is the higher proportion of nonagentive causer (NC) in causative uses of verbs that are used more frequently as a noncausative. Nonagentive causers are less frequent in verbs that are predominantly used as a causative.

Taking all 12 verbs together, the overall pattern is that the ease of cause identification correlates with frequency of verb uses: we see that verbs that are predominantly used as a causative tend to have a higher percentage of agent causer, a cause type with the highest degree of cause identifiability. To further highlight the nature of this correlation, let us now continue on to present smoothed scatter plots which show the relationship between the percentage of verb uses and cause type prominence.

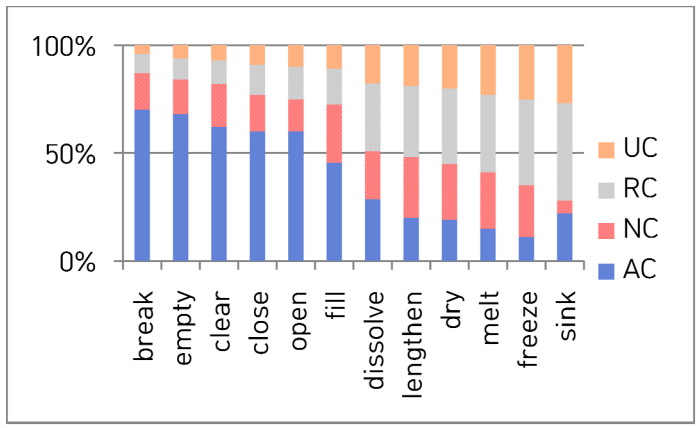

Figure 3. Proportion of cause type

In Figures 4 and 5, the x-axis shows the agent causer prominence scale, that is, the rank of the verbs obtained by calculating the ratio of agent causer in total use of each verb (= AC/AC + NC + RC + UC); the y-axis in Figure 4 shows the percentage of the causative use of each verb, and the y-axis in Figure 5 shows the percentage of the noncausative use.[8]

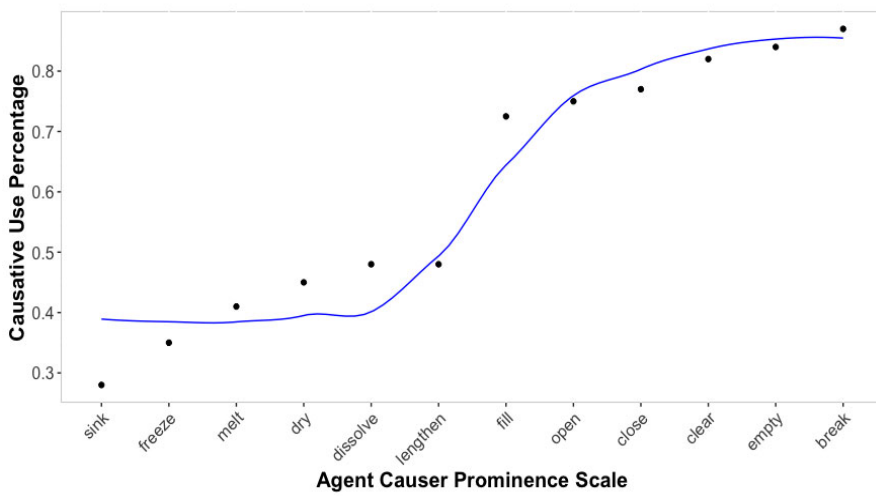

Figure 4. Smoothed scatter plot of the relationship between the percentage of causative use and the agent causer prominence scale

8 All plots and statistical analyses were obtained by R version 4.4.1 (R Core Team 2021).

Figure 4 shows at first sight that there is a positive correlation between the percentage of the verbs' causative use (%C) and their agent causer prominence. Crucially, there is a clear tendency for the percentage of the causative use to increase with the increase in agent causer prominence: the trendline is going up. Conversely, we find a negative correlation between the percentage of the verbs' noncausative use (%N) and their agent causer prominence. As shown in Figure 5, the trendline is going down.

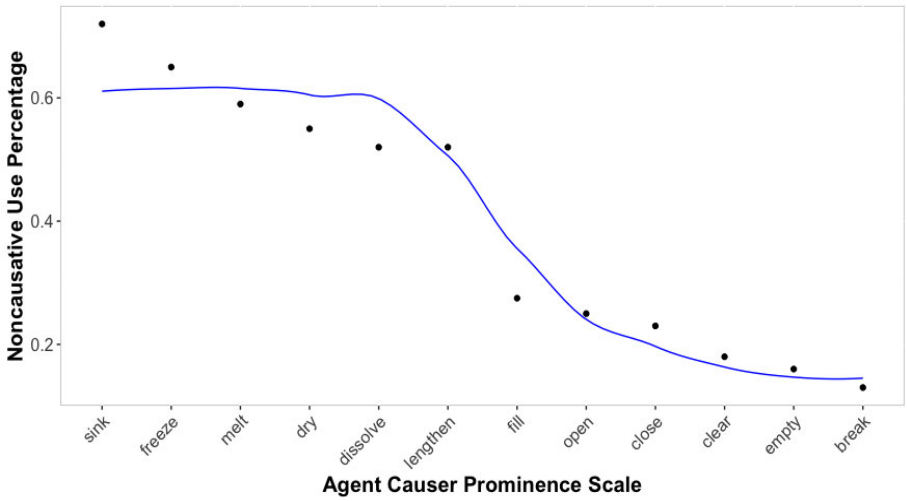

Figure 5. Smoothed scatter plot of the relationship between the percentage of noncausative use and the agent causer prominence scale

In order to see whether the trends observed were statistically significant, a series of the Kendall Tau (τ) Rank test have been performed, as detailed below. This non-parametric correlation test provides a Tau (τ) value, which ranges from −1 (perfect negative correlation) to +1 (perfect positive correlation). and an associated *p*-value indicating how sure one can be of its significance.[9] The results are given in Table 3.

9 There are two widely used non-parametric rank correlation tests: the Kendall Tau (τ) Rank test and the Spearman Rho (ρ) Rank test. Although in most situations, the interpretations of Kendall's Tau and Spearman's Rho correlation coefficient are very similar and thus lead to the same inferences, the Kendall's Tau should be preferred if one has a small dataset and many ranks are tied (Field et al. 2012: 225). This study reports Kendall's Tau as using it has the following advantages: *p*-values associated with it are more accurate with smaller sample sizes, and the interpretation of it in terms of the pairs of ranks is very direct.

Table 3. Results of the Kendall's Tau Rank test

Variables	Kendall's Tau (τ)	*p*-value (two-sided)
%C – AC prominence	0.879	< 0.0001
%N – AC prominence	−0.848	0.0001
%C – RC + UC prominence	−1.000	< 0.0001
%N – RC + UC prominence	0.970	< 0.0001

As shown in Table 3, a Kendall Tau Rank test correlation of the percentage of verbs' causative use against their agent causer prominence shows that the correlation between these two variables has a positive Tau value, i.e., 0.879, indicating a strong positive trend, and a very high associated *p*-value ($p < 0.0001$). Conversely, a Kendall Tau Rank test correlation of the percentage of verbs' noncausative use against their agent causer prominence shows that the Tau value amounts to −0.848, indicating a strong negative trend, and has a highly significant *p*-value ($p = 0.0001$).

Figures 6 and 7 are plots which show the relationship between the percentage of verb uses and the scale of recoverable and unknown cause prominence. In these figures, the x-axis shows the rank of the verbs obtained by calculating the ratio of the sum of recoverable cause and unknown cause in total use of each verb (= RC + UC/AC + NC + RC + UC); the y-axis in Figure 6 shows the percentage of the causative use of each verb, and the y-axis in Figure 7 shows the percentage of the noncausative use.

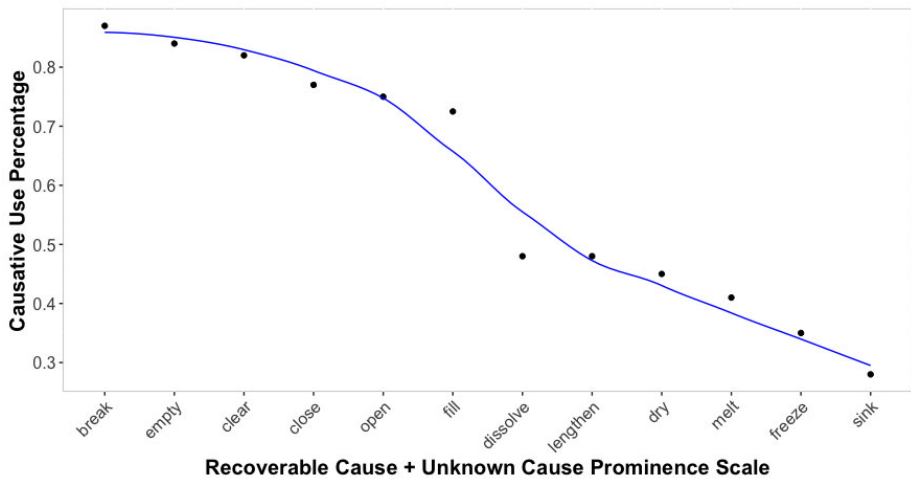

Figure 6. Smoothed scatter plot of the relationship between the percentage of causative use and the RC+UC prominence scale

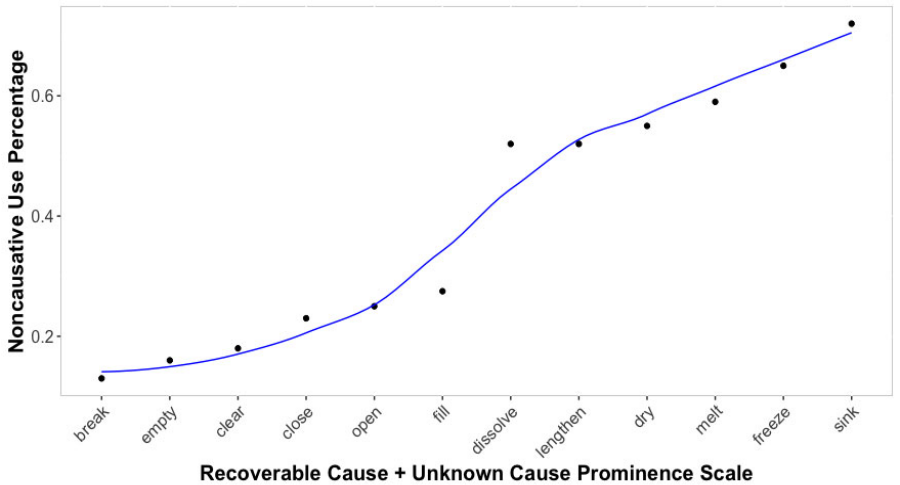

Figure 7. Smoothed scatter plot of the relationship between the percentage of noncausative use and the RC+UC prominence scale

We observe a negative correlation between the percentage of the verbs' causative use and their recoverable and unknown cause prominence. As shown in Figure 6, the percentage of the verbs' causative use tends to decrease with the increase in recoverable and unknown cause prominence. The Tau value which amounts to −1 ($p < 0.001$) confirms this impression and indicates a robust negative correlation between the two variables (see the table in (22)). Conversely, Figure 7 shows that there is a positive correlation between the percentage of the verbs' noncausative use and their recoverable and unknown cause prominence. This impression is confirmed by the calculated Tau value 0.970 ($p < 0.0001$), indicating a strong positive correlation between the two variables. Thus, these results confirm the predictions tested in this study that cause identifiability and frequency of verb uses covary in different ways – a high degree of cause identifiability increases the likelihood that the causative is used, and a low degree of cause identifiability increases the likelihood that the noncausative is used.

In summary, on the basis of analyses of the BNC data, this section has shown that the ease of cause identification correlates significantly with frequency of verb uses. In the next section, we will zoom in on frequency variation in the causative alternation, and discuss its explanations.

5. General discussion

In addition to showing a clear division of labor between the two variants in the expression of cause type, the corpus data also show evidence for lexical variation in the causative alternation. In this paper, I propose that this variability in verb-specific preferences for one variant over the other is driven by the usage probability of occurrence of an identifiable cause: verbs that are more likely to describe eventualities that have a clear identifiable cause tend to be more frequently used as a causative while verbs that are more likely to describe eventualities that have a less clear cause tend to be more frequently used as a noncausative. Thus, cause identifiability as a contextual property of eventualities underlies both lexical and grammatical variation in the causative alternation as well as the correlation between cause type and verb uses.

Language structure and use are ultimately grounded in the communicative needs of language users, and the observed correlation between cause identifiability and verb uses, too, is likely due to the tendency for languages to use structure efficiently. Levshina (2022) proposes that language users' behavior is guided by several principles, which explain how they can behave efficiently in everyday communication. The three main principles are as follows: (i) the principle of positive correlation between benefits and costs, (ii) the principle of negative correlation between accessibility and costs, and (iii) the principle of maximization of accessibility. The correlation between cause identifiability and verb uses can be interpreted as being a consequence of the first principle. According to this principle, high costs are associated with high benefits, and low costs are associated with low benefits. This means that language users should spend more effort and time on information that provides more benefits, and less effort and time on less useful information (Levshina 2022: 22).[10] From the efficiency perspective, the causative variant, which specifies the cause of a change of state, is more costly, and it is at the same time more informative than the noncausative variant, which expresses just the change of state. As Rappaport Hovav (2014) pointed out, the choice of the causative variant is preferrable unless the extra information, i.e., the cause of the change of state, is redundant, unimportant or unidentifiable. Therefore, the high costs required for

10 This principle is closely related to Grice's Maxims of Quantity, 'Make your contribution as informaive as required (for the current purposes of the exchange)', and 'Do not make your contribution more informative than is required' (Grice 1975: 45), as well as Horn's (1984) Q-principle: 'say as much as you can', and the R-principle: 'say no more than you must'.

producing a more complex structure are justified by the high relevance and importance of the identifiable agent in the description of a change of state and the high informativeness of the causative variant. These are the benefits of communication in a very broad sense, and ultimately motivate the efficient division of labor between the two causative alternation variants.

An alternative way to account for the frequency variation in the causative alternation would be to suggest that it is driven by a lexical property of verbs, namely verbs' spontaneity. A verb's spontaneity is a lexical semantic property that can be described in terms of the type of event it denotes: spontaneous verbs denote events without a salient external cause, while non-spontaneous verbs denote events with a salient external cause. In a recent corpus study, Samardžić and Merlo (2018) revisited the notion of external causation. They define the probability of external causation as the degree of involvement of an external causer in an event described by a verb. They call this property *Sp* for spontaneity (following Haspelmath 1993) and take the degree of *Sp* to be a general scalar component in the lexical representation of the alternating verbs, assigned as a property to the verb type that is common to the causative and noncausative realizations (Samardžić and Merlo 2018: 905). Statistical analyses of data extracted from the English side of the parallel corpus Europarl indicate that the corpus-based measure of *Sp* is correlated with the rank of the verbs based on the C/A ratio calculated on the typology of the morphological marking of the verbs across languages.[11] Samardžić and Merlo (2018) take this correlation as evidence suggesting that the probability of external causation is a grammatically relevant lexical property that influences both the corpus frequency distribution of structural realizations of alternating verbs and typological distribution of morphological marking.

Another interesting finding of this work is that the *Sp*-value is distributed normally over a large set of verbs listed by Levin (1993). Since the normal distribution is symmetric and concentrated around the mean, this finding suggests that most alternating verbs can be expected to describe events in which there is around 50 percent chance for an external causer to occur. In fact, Samardžić and Merlo's (2018: 914) analysis shows that most verbs are neutral with respect to external causality, whereas only some tend

11 The *Sp*-value for each verb *v* is calculated as the logarithm of the C/A ratio, namely the ratio between the proportions of causative and anticausative (a term used to refer to noncausative by Samardžić and Merlo (2018)) uses of the verb in the corpus. See Samardžić and Merlo's (2018: 912) for more detailed discussion of estimation and interpretation of the *Sp*-values.

to be assigned extreme *Sp*-values. If it turns out that the great majority of alternating verbs are the ones that describe events whose external causer can be present or absent with similar probability, this raises the following questions that are not addressed by Samardžić and Merlo (2018): what determines the appearance of the cause argument in a given context? Why one variant is chosen rather than the other when, in principle, both might be possible? Is the choice a matter of lexical, semantic, discourse or contextual considerations? The results of the corpus analysis presented in this study offer some answers to these questions, providing new empirical support for the view that contextual factors play an important role in this choice, as originally suggested by Rappaport Hovav (2014).

Further, the corpus data indicate that spontaneity or the involvement of an external causer is insufficient to account for the frequencies of the causative and noncausative uses of alternating verbs. This can be shown by examining corpus frequencies of natural changes of state, namely changes of state that come about in the natural order events without the intervention of an agent or without a considerable input of external force or energy. The meaning of the investigated verbs is compatible with describing either natural or non-natural (externally caused) changes of state although they may differ in the probability of description of kinds of change of state. But since there is no independent way of measuring the naturalness of a change, it was determined by looking at choice of patients found with the verbs. For instance, examples where the verb *melt* is found with ice or ice cream or other substances that melt at room temperature have been counted as descriptions of a natural change of state; examples where *melt* is found with substances that do not melt at room temperature have been counted as descriptions of a non-natural change of state.

Tables 4 and 5 show the counts for uses describing natural changes of state of Set A verbs and Set B verbs, respectively. The numbers in parentheses show the percentage of verb uses describing natural changes of state in the total number of examples of each verb (200 examples per verb). As shown in the tables, descriptions of a natural change of state are more frequent in uses of verbs that are more frequently used as a noncausative. In particular, over 20 percent of the uses of three verbs – *dry, melt* and *freeze* – were descriptions of a natural change of state. The percentage for other verbs was much lower: 10 percent for *dissolve,* and only 5 percent for *sink* and *lengthen.* The percent for *break, clear, empty*, and *open* was even lower, and none of the uses of *close* and *fill* in the data were descriptions of a natural change of state.

Table 4. Frequency of descriptions of a natural change of state (Set A)

dissolve	dry	fill	freeze	melt	sink
20 (10 %)	46 (23 %)	0 (0 %)	48 (24 %)	50 (25 %)	10 (5 %)

Table 5. Frequency of descriptions of a natural change of state (Set B)

break	clear	close	empty	lengthen	open
3 (1.5 %)	6 (3 %)	0 (0 %)	2 (1 %)	10 (5 %)	4 (2 %)

All instances of descriptions of a natural change of state in the data were realized as a noncausative. Examples of noncausative uses of the investigated verbs describing natural events from the BNC are illustrated in (20) and (21).

(20) Set A verbs

a. The sugars dissolve into the water and the sweet liquid, called wort, is pumped to a copper. (BNC miscellaneous texts)

b. Sweat began to dry and strength seeped back into my limbs. (BNC miscellaneous texts)

c. In 1947, when the canal froze, you could skate from here to Abingdon --- nearly ten miles. (BNC miscellaneous texts)

d. When the ice melted, some shallow lakes remained where boulder clay blocked old river courses. (BNC non-academic: social sciences)

e. The muffled voice behind the door sank at last into hopeless silence. (BNC fiction: prose)

(21) Set B verbs

a. Dawn was breaking when they reached Camborne and Celia stirred a little. (BNC fiction: prose)

b. Early fog had cleared and the airport manager and I were standing on the tarmac lining up the motorcade when my car phone rang. (BNC commerce, finance, economics)

c. Cranston remained full of good cheer, aided and abetted by an apparently miraculous wineskin which never seemed to empty. (BNC fiction: prose)

d. Some, such as carnations, radishes, scarlet pimpernels and clover, flower in the spring when the days are lengthening, and so are called long-day plants. (BNC non-academic: natural sciences)
e. For a few moments the clouds open and we lie in the heat of a lazy sun, our efforts just rewarded. (BNC miscellaneous texts)

Importantly, the majority of noncausative uses of the investigated verbs were descriptions of events involving an external cause that is difficult to identify. The overall high frequency of noncausative uses of externally caused changes of state indicates that spontaneity or the involvement of an external causer is not a major concern in choosing the noncausative variant over the causative. Rather, the results are in line with the proposal that the noncausative is chosen because it allows for non-expression of a cause that is unimportant, known or difficult to identify, thus favoring the explanation based on cause identifiability.

To sum up, this section has discussed evidence based on corpus frequencies of natural changes of state which supports the claim that cause identifiability is a crucial factor that accounts for the frequencies of the causative and noncausative uses of alternating verbs. Other lexical semantic factors discussed in the literature, that is, spontaneity or the involvement of an external causer, are less successful in explaining the corpus frequency distribution of syntactic realizations of alternating verbs and the choice of a variant in a given context.

6. Conclusion

This paper has investigated the frequency variation in the English causative alternation. The main questions were why verbs differ in their frequency of use in the two variants and what contextual factors explain verb-specific preferences for one variant over the other. On the basis of a corpus study of 2400 instances of causative and noncausative uses of 12 alternating verbs, I have shown that the ease of cause identification correlates significantly with frequency of verb uses. Based on this evidence, I have argued that cause identifiability is a crucial factor that accounts for variation in the English causative alternation.

The present study is the first to my knowledge to demonstrate applying computational

corpus linguistic methodology that cause identifiability better accounts for the corpus frequency distribution of the alternation variants and verb-specific preferences for one variant over the other than other lexical semantic factors discussed in the literature. This work can be considered as a basis for further investigation of the realization of causal events, which will extend the empirical validation of the hypothesis tested here to a larger set of verbs and verb uses from different corpora. Evidence from a large-scale analysis will contribute to uncovering the nature of native speakers' knowledge of the complex probabilistic constraints mediating the realizations of causal events and event participants.

References

Alexiadou, Artemis, Elena Anagnostopoulou, and Florian Schäfer. 2006. The properties of anti-causatives crosslinguistically. In Mara Frascarelli (ed.), *Phases of interpretation,* 175-199. Berlin; New York: Mouton de Gruyter.

Christ, Oliver. 1994. A modular and flexible architecture for an integrated corpus query system. *Proceedings of COMPLEX '94*, 23-32. Budapest.

Comrie, Bernard. 2006. Transitivity pairs, markedness, and diachronic stability. *Linguistics* 44(2): 303-318.

Croft, William. 1991. *Syntactic categories and grammatical relations: The cognitive organization of information*. Chicago, IL: University of Chicago Press.

Dixon, Robert M.W. 2000. A typology of causatives: Form, syntax and meaning. In Robert M.W. Dixon and Alexandra Y. Aikhenvald (eds.), *Changing valency: Case studies in transitivity,* 30-83. New York: Cambridge University Press.

Evert, Stefan and Andrew Hardie. 2011. Twenty-first century corpus workbench: Updating a query architecture for the new millennium. *Proceedings of the Corpus Linguistics 2011 Conference.* University of Birmingham, UK.

Field, Andy, Jeremy Miles, and Zoë Field. 2012. *Discovering statistics using R*. Los Angeles, CA: Sage.

Fodor, Jerry A. 1970. Three reasons for not deriving 'kill' from 'cause to die', *Linguistic Inquiry* 1(4): 429-438.

Grice, H. Paul. 1975. Logic and conversation. In Peter Cole and Jerry L. Morgan (eds.), *Syntax and semantics, Volume 3: Speech acts,* 41-58. New York: Academic Press.

Hardie, Andrew. 2012. CQPweb – combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics* 17(3): 380-409.

Haspelmath, Martin. 1993. More on typology of inchoative/causative verb alternations. In Bernard Comrie and Maria Polinsky (eds.), *Causatives and transtivity, Volume 23,* 87-121. Amsterdam;

Philadelphia: John Benjamins Publishing Company.
Haspelmath, Martin. 2008. Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics* 19(1): 1-33.
Haspelmath, Martin. 2016. Universals of causative and anticausative verb formation and the spontaneity scale. *Lingua Posnaniensis* 58(2): 33-63.
Haspelmath, Martin, Andreea Calude, Michael Spagnol, Heiko Narrog, and Elif Bamyaci. 2014. Coding causal-noncausal verb alternations: A form-frequency correspondence explanation. *Journal of Linguistics* 50(3): 587-625.
Heidinger, Steffen. 2015. Causalness and the encoding of the causative-anticausative alternation in French and Spanish. *Journal of Linguistics* 51(3): 562-594.
Hoffmann, Sebastian, Stefan Evert, Nicholas Smith, David Y. W. Lee, and Ylva Prytz. 2008. *Corpus linguistics with BNCweb – A practical guide*. Frankfurt am Main: Peter Lang.
Horn, Laurence R. 1984. Towards a new taxonomy for pragmatic inference: Q-based and R-based Implicature. In Deborah Schiffrin (ed.), *Georgetown University Round Table on languages and linguistics*, 11-42. Washington, DC: Georgetown University Press.
Koontz-Garboden, Andrew. 2007. *States, changes of states, and the monotonicity hypothesis*. PhD Dissertation. Stanford University.
Koontz-Garboden, Andrew. 2009. Anticausativization. *Natural Language & Linguistic Theory* 27(1): 77-138.
Lakoff, George. 1990. *Women, fire and dangerous things*. Chicago, IL: University of Chicago Press.
Lee, Minhaeng. 2015. *Language research in the era of big data* [In Korean]. Seoul: The 21st Century Books.
Lee, Minhaeng. 2021. *A computational corpus linguistic study of the dependency grammar of German* [In Korean]. Seoul: Yeklak.
Levin, Beth. 1993. *English verb classes and alternations*. Cambridge, MA: The MIT Press.
Levin, Beth. 2015. Semantics and pragmatics of argument alternations. *The Annual Review of Linguistics* 1(1): 63-83.
Levin, Beth and Malka Rappaport Hovav. 1995. *Unaccusativity: At the syntax-lexical semantics interface*. Cambridge, MA: The MIT Press.
Levshina, Natalia. 2022. *Communicative efficiency: Language structure and use*. Cambridge: Cambridge University Press.
Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford Core NLP natural language toolkit. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations,* 55-60.
de Marneffe, Marie-Catherine and Christopher D. Manning. 2008/2016. Stanford typed dependencies manual. Technical report, Stanford University.
McCawley, James D. 1978. Conversational implicature and the lexicon. In Peter Cole (ed.), *Syntax and semantics, Volume 9: Pragmatics*, 245-259. New York: Academic Press.

McKoon, Gail and Talke Macfarland. 2000. Externally and internally caused change of state verbs. *Language* 76(4): 833-858.

Nedyalkov, Vladimir P. and G. G. Silnitsky. 1973. The typology of morphological and lexical causatives. In Ferenc Kiefer (ed.), *Trends in Soviet theoretical linguistics*, 1-32. Dordrecht: Springer.

Nichols, Johanna, David A. Peterson, and Jonathan Barnes. 2004. Transitivizing and detransitivizing languages. *Linguistic Typology* 8(2): 149-211.

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

Qi, Peng, Yuhao Zhang, YuHui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A python natural language toolkit for many human languages. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations,* 101-108.

Rappaport Hovav, Malka. 2014. Lexical content and context: The causative alternation in English revisited. *Lingua* 141: 8-29.

Rappaport Hovav, Malka and Beth Levin. 2012. Lexicon uniformity and the causative alternation. In Martin Everaert, Marijana Marelj, and Tal Siloni (eds.), *The theta system: Argument structure at the interface,* 150-176. Oxford: Oxford University Press.

R Core Team. 2021. The R project for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Reinhart, Tanya. 2002. The theta system - An overview. *Theoretical Linguistics* 28(3): 229-290.

Reinhart, Tanya. 2016. *Concepts, syntax and their interface*. Cambridge, MA: The MIT Press.

Samardžić, Tanja and Paola Merlo. 2018. The probability of external causation: An empirical account of crosslinguistic variation in lexical causatives. *Linguistics* 56(5): 895-938.

Schäfer, Florian. 2008. *The syntax of (anti-)causatives: External arguments in change-of-state contexts*. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Shibatani, Masayoshi. 1976. The grammar of causative constructions: A conspectus. In Masayoshi Shibatani (ed.), *Syntax and semantics, Volume 6: The grammar of causative constructions,* 1-40. New York: Academic Press.

Smith Carlota S. 1978. Jespersen's 'move and change' class and causative verbs in English. In Moḥammed Alī Jazāyerī, Edgar C. Polomé, and Werner Winter (eds.), *Linguistic and literary studies in honor of Archibald A. Hill, Volume 2: Descriptive linguistics,* 101-109. The Hague: Mouton.

Talmy, Leonard. 1976. Semantic causative types. In Masayoshi Shibatani (ed.), *Syntax and semantics, Volume 6: The grammar of causative constructions,* 43-116. New York: Academic Press.

Talmy, Leonard. 2000. *Toward a cognitive semantics*. Cambridge, MA: The MIT Press.

Thompson, Sandra A. 1987. The passive in English: A discourse perspective. In Robert Channon and Linda Shockey (eds.), *In honor of Ilse Lehiste: Ilse Lehiste puhendusteos,* 497-511. Dordrecht: Foris.

Wanner, Anja. 2009. *Deconstructing the English passive*. Berlin; New York: Mouton de Gruyter.

Wolff, Phillip. 2003. Direct causation in the linguistic coding and individuation of causal events. *Cognition* 88(1): 1-48.

Wright, Saundra. 2001. *Internally caused and externally caused change of state verbs.* PhD Dissertation. Northwestern University.

Wright, Saundra. 2002. Transitivity and change of state verbs. *Proceedings of the 28th Meeting of the Berkeley Linguistics Society,* 339-350. Berkeley: Berkeley Linguistics Society.

Hanjung Lee
Professor
Department of English Language and Literature
Sungkyunkwan University
25-2 Sungkyunkwan-ro, Jongro-gu,
Seoul 03063, Korea
E-mail: hanjung@skku.edu

Received: 2023. 10. 11.
Revised: 2023. 11. 17.
Accepted: 2023. 11. 17.