16

Covarying Collexemes in the *Into*-causative

STEFAN TH. GRIES AND ANATOL STEFANOWITSCH*

1 Introduction

In investigations of how words and constructions interact, the focus is typically on the preferences or restrictions associated with one particular slot in the construction; little attention is paid to possible interactions between two or more such slots (see Rohde 2001 for a rare exception). However, such interactions are intuitively important at least for some constructions, for example the one exemplified in (1):

(1) a. He *tricked* me into *employing* him.
b. His aim was to *force* the Government into *holding* a plebiscite.
c. He had been *coerced* and *terrorized* into *absconding*.

This construction, which we call *into*-causative, has two predicates: the main verb (which encodes a causing event, and will be called the *cause predicate*), and the present participle in the oblique introduced by *into* (which encodes a result, and will be called the *result predicate*).

In this paper we extend a 'single-slot' methodology developed in Stefanowitsch and Gries (2003) to the investigation of potential interactions between two slots and apply it to the *into*-causative. We show that such interactions exist, i.e. that cause and result predicates 'covary' systematically. We then consider two factors influencing this covariation: a cognitive one, based in causative event types; and a cultural one, based in knowledge about frames and possible cause-effect relations between them.

*The order of authors is arbitrary. The authors would like to thank Britta Mondorf and André Schäfer for supplying the raw data from *The Guardian* used in this study.

Language, Culture, and Mind.
Michel Achard and Suzanne Kemmer (eds.).

2 Corpora and Constructions

2.1 Collexemes

A recent corpus-based study of the *into*-causative (Hunston and Francis 2000) provides some frequency data for the cause-predicate slot, noting a tendency of the construction to occur with verbs of negative emotions (e.g. *frighten*, *intimidate*, *panic*, *scare*, *terrify, embarrass, shock, shame,* etc.) or ways of speaking cleverly/deviously (e.g. *talk, coax*, *cajole*, *charm*, *browbeat,* etc.); in addition, the authors claim that the construction is associated with 'some kind of forcefulness or even coercion' (Hunston and Francis 2000: 106). Unfortunately, this analysis is not particularly systematic (for example, two of the most frequent verbs in this construction, *force* and *coerce*, are not even mentioned); moreover, their analysis lacks a quantitative assessment of the association strength between the verbs and the construction.

In a previous study (Stefanowitsch and Gries 2003) we developed a method for assessing the association strength between words and constructions (which we refer to as *collostructional analysis*), and applied it to the *into*-causative (among other constructions). We calculated association strengths on the basis of a crosstabulation of the individual frequencies of the word and the construction in question as well as their joint frequency, using the Fisher exact test (which, unlike other measures sometimes suggested, does not make unwarranted assumptions about the distribution of natural language data and works even with low frequencies). For example, to calculate the association strength (in our terms, *collostruction strength*) between the verb *talk* and the *into*-causative, the data in Table 1 is needed.

	talk	other verbs	Total
into-causative	62	1,524	1,586
other constructions	28,637	10,176,077	10,204,714
Totals	28,699	10,177,601	10,206,300

Table 1. Frequencies of *talk* and the *into*-causative in the *BNC*

The Fisher exact p-value for this distribution is 2.38E-48, indicating that the association between *talk* and the cause-predicate slot of the *into*-causative is highly significant. Once this calculation is performed for all verbs occurring in a given slot, their collostruction-strength ranking can be determined by sorting them by their p-value (see Stefanowitsch and Gries 2003 for justification). Table 2 lists the twenty cause predicates most strongly associated with the *into*-causative (we refer to words significantly associated with a particular slot in a construction as *collexemes*).

CAUSE-PRED. COLLEXEME	COLLOSTRUCTION STRENGTH	CAUSE-PRED. COLLEXEME	COLLOSTRUCTION STRENGTH
1. *trick*	2.11E-267	11. *pressure*	3.88E-85
2. *fool*	1.68E-187	12. *cajole*	4.08E-85
3. *coerce*	1.15E-158	13. *blackmail*	3.31E-64
4. *force*	6.31E-136	14. *dupe*	7.77E-52
5. *mislead*	9.57E-110	15. *coax*	6.00E-51
6. *bully*	2.53E-109	16. *delude*	8.83E-49
7. *deceive*	5.94E-109	17. *talk*	2.38E-48
8. *con*	4.41E-102	18. *goad*	1.35E-46
9. *pressurize*	4.80E-101	19. *shame*	1.28E-45
10. *provoke*	4.05E-87	20. *brainwash*	2.42E-37

Table 2. Strongest collexemes for the *into*-causative (*BNC*)

The verb most strongly associated with the construction is *trick*, followed by *fool, coerce,* and *force*. As is immediately obvious, these results differ strongly from the frequency counts presented by Hunston and Francis—a much more complex picture emerges. There seem to be at least three strongly entrenched classes of verbs in this construction: verbs like first- and second-ranked *trick*, *fool,* as well as *mislead* instantiate a sense that might be characterized as TRICKERY; *coerce*, *force*, *bully*, etc. instantiate a sense one might describe as PHYSICAL FORCE; *blackmail*, *cajole*, *coax*, *goad,* and *shame* form a group that might be described as VERBAL PERSUASION. Each of these three groups is instantiated by several verbs among the top twenty collexemes, and collostructional analysis suggests that the TRICKERY sense is the one that is most strongly entrenched. Verbs of negative emotion, in contrast, are only instantiated once among the top twenty collexemes (*shame*).

There is no room here for a detailed analysis of these facts or an attempt to capture the invariant semantics of the *into*-causative (cf. Wierzbicka 1998 for a partial attempt), or to show how the different senses are related. Suffice it to draw attention to one interesting phenomenon which emerged from our previous study: different verbs as well as different classes of verbs in the cause slot of the *into*-causative seem to have different restrictions or preferences with respect to the result predicates they cooccur with. For example, first-ranked *trick* seem much less restricted than other verbs from the same semantic group with respect to its result predicates. Also, it seems that verbs from the PHYSICAL FORCE group exhibit a tendency to occur with action verbs, but not with verbs of perception and cognition. We did not investigate this intraconstructional covariation of cause and result predicates

systematically in our previous study, but the results are intriguing enough to extend collostructional analysis to this kind of phenomenon.

2.2 Covarying Collexemes

Extending collostructional analysis to the issue of covarying collexemes essentially involves comparing the collostruction strength of the lexemes in the two slots in relation to each other; we must determine for each word occurring in one of the slots ($collexeme_1$) which words in the other slot ($collexeme_2$) cooccur with it significantly more often than expected.

First, this requires us to find all instances of the construction in question and then determine the frequency of each pair of potential collexemes. We extracted all cases of the *into*-causative from the 1990-2000 volumes of *The Guardian*; this yielded 6,288 tokens. We then identified the verb lemmas in the cause- and result-predicate slots, yielding 3,908 combinations (some examples are shown in Table 3).

$COLLEXEME_1$-$COLLEXEME_2$	N	$COLLEXEME_1$-$COLLEXEME_2$	N
force - make	51	*coerce - do*	8
bully - accept	20	*fool - see*	6
push - accept	15	*lure - vote*	2

Table 3. Selected collexeme-combination frequencies in the *into*-causative

It would clearly not do to simply assume that the frequent combinations are also the important combinations, i.e. the ones that will allow us to gain insights into the covariation of collexemes in the construction in question. Instead, it has to be determined which of these combinations occur more frequently than expected given the separate overall frequencies with which $collexeme_1$ and $collexeme_2$ occur in the construction in question. For example, *push into accepting* occurs fifteen times, which makes *push-accept* the thirteenth most frequent combination, and therefore potentially important under a simple frequency approach. However, *push* and *accept* occur in the *into*-causative as $collexeme_1$ and $collexeme_2$ 295 and 296 times respectively. Given 6,288 *into*-causatives, we would thus expect 13.88 (i.e. (295×296)/6,288) occurrences of *push into accepting* just by chance. Obviously, the observed frequency of fifteen is not significantly higher than what would be expected on a chance level (Fisher exact, p=0.417), demonstrating that *push into accepting* is not among the configurations to which much attention should be directed.

In sum, we need an adjustment that takes into consideration the frequencies of the two potential collexemes in the same way that measures of collocational strength generally adjust for expected frequencies (cf., e.g.,

Church and Hanks 1990, Church et al. 1991, Dunning 1993). Extending arguments in Gries (2003), we again propose to use the Fisher exact test to make this adjustment for the same reasons mentioned above.

Thus, in order to identify all instances of significant attraction between verbs in the cause- and result-predicate slots, we computed expected frequencies and Fisher exact p-values for all 3,908 combinations types yielding 1,487 significant combinations. The most interesting results of this analysis will be presented in the next section.

3 Covarying Collexemes in the *Into*-causative: Cognitive and Cultural Factors

3.1 Data

In order to reduce the complexity of the data somewhat, we decided to discard from initial consideration all combinations occurring only once, as well as all combinations occurring *less* frequently than expected (although the latter may well be interesting, cf. Stefanowitsch and Gries 2003). This left us with 361 combination types, accounting for 1,478 tokens out of the original 6,288 tokens. The 25 most strongly attracted combinations are shown in Table 4.

Most striking about these pairs is that many of them seem to be based in frame-semantic knowledge of varying degrees of culture-specificity about what is likely to cause what, i.e. what frames stand in an entrenched cause-effect relationship in a particular culture. Clear examples are first-ranked *bounce-accept*, as well as *torture-confess*, *terrorize-flee*, and, of course, *con-pay* and *mislead-buy,* which no doubt are very culture-specific.

Other combinations are also of this type although they require more context to be interpretable. For example, *dragoon-vote* seems strange given a culture with a strong democratic tradition like Great Britain, but once you take a closer look it turns out that the examples mostly describe situations where party members are being convinced to vote in the interest of party discipline for something they do not believe in, or even something detrimental to them.

Yet other combinations have light verbs as result predicates; here a closer look at the light-verb constructions involved may also be revealing. For example, *force-make* (incidentally the most frequent combination) and *rush-make* are both significant combinations, but in the case of *force* a wide range of result predicates are involved, including *make a decision*, *statement*, *mistake*, *concession*, and so on. In contrast, *rush* occurs with *make a decision* in the majority of cases (seven out of eleven); this suggests that the

RUSH and DECIDE frames are in a culturally entrenched relationship in British culture.

COLLEXEME$_1$ *Cause predicate*	COLLEXEME$_2$ *Result predicate*	FREQUENCY	$P_{Fisher\text{-}Exact}$
bounce	*accept*	29	8.442E-15
torture	*confess*	8	7.001E-14
draw	*comment*	6	2.625E-11
shock	*understand*	7	3.287E-11
stimulate	*produce*	6	4.678E-10
dupe	*carry*	8	5.698E-08
con	*pay*	16	9.571E-08
hoodwink	*leave*	8	1.042E-07
mislead	*buy*	14	1.048E-07
delude	*suppose*	3	1.616E-07
terrorize	*flee*	4	1.729E-07
talk	*let*	12	1.805E-07
dupe	*leave*	13	2.461E-07
force	*make*	51	2.844E-07
pressure	*have*	14	3.128E-07
bounce	*announce*	6	7.937E-07
shame	*cleaning*	4	1.116E-06
dragoon	*vote*	7	1.263E-06
swing	*plan*	2	3.034E-06
fool	*queue*	3	3.672E-06
lock	*use*	5	3.927E-06
guide	*lend*	2	4.247E-06
rush	*make*	11	4.95E-06
educate	*understand*	3	5.059E-06
fool	*see*	6	6.602E-06

Table 4. Most significant collexeme combinations in the *into*-causative

For other verbs, an explanation in terms of cultural frames is less immediately plausible. For example, *shock-understand* does not encode a combination of frames that would be regarded as typically standing in a cause-effect relationship. Here, a look at the other three cognition verbs in the result-predicate slot (*delude-suppose*, *educate-understand*, *fool-see*) shows an interesting asymmetry that may point to a different type of influencing factor: it seems that cognition verbs as result predicates prefer cognition verbs as cause predicates, while action-verb results take either cognition verbs (e.g. *dupe-carry* or *con-pay*) or action verbs (e.g. *bounce-announce*, *guide-lend*). In these cases a cognitive explanation based on general causative

event types may be more helpful. In the next section we will briefly discuss such an explanation before returning to the frame-semantic approach.

3.2 Cognitive Factors

In an analysis of analytic causative constructions, Stefanowitsch (2001) proposes three causative event types that are argued to reflect preferred construals of causal chains. First, the MANIPULATE type, where an animate causer intentionally acts on a causee in a way that influences the causee in such a way that he or she performs some activity (e.g. *They were tortured into confessing*). Second, the TRIGGER type, where an event occurs which influences a causee in such a way that, given the nature of the causee, the causee will inevitably undergo some process (e.g. *Drunken drivers will be shocked into understanding the error of their ways* in the mental domain, and *The vaccines could stimulate the immune system into producing defense cells* in the biological domain). Third, the PROMPT type, where an event occurs and a causee perceives this event and decides to react by performing some activity (e.g. *The government was shamed into cleaning up its stock market*).

If these event types indeed reflect preferred construals of causative events, we might expect the following distribution of cause and effect predicates in the *into*-causative: action verbs as result predicates should be able to occur equally well with action verbs as cause predicates (reflecting the MANIPULATE type) and with cognition verbs as cause predicates (reflecting the PROMPT type). In contrast, cognition verbs as result predicates should prefer cognition verbs as cause predicates (reflecting the TRIGGER type), and disprefer action verbs (as this combination does not reflect any of the three causative-event types). Table 5 shows the observed and expected co-occurrence frequencies of action and cognition verbs.

		RESULT PREDICATE		
		Action	Experiencer	Total
CAUSE PREDICATE	Action	Observed 4324 Expected 4304	Observed 96 Expected 116	4420
	Experiencer	Observed 1788 Expected 1808	Observed 69 Expected 49	1857
	Total	6112	165	6277

Table 5. Action and experiencer verbs in the *into*-causative

First of all, note that there is a drastic difference in the proportion of mental and physical causes, with physical causes more than twice as fre-

quent as mental causes. In the case of results, this difference is even more pronounced; action verbs are thirty-seven times more frequent than cognition verbs. This in itself is an interesting result; it may be a manifestation of a cognitive preference to perceive and conceptualize causation in the physical domain.

With respect to the issue of preferred causative event types, the data in Table 3 show that there is indeed a weak but significant influence of the type of preference hypothesized above: taking into account the general preference for physical causes and results, physical results are equally strongly associated with physical or mental causes (action-action χ^2=0.09, n.s., action-experience χ^2=0.22, n.s.), but mental causes occur more frequently than expected with mental results (χ^2=8.35, ***), and less frequently with physical ones (χ^2=3.51, n.s.).

3.3 Cultural Frames

We will now take a closer look at some combinations plausibly based on knowledge about culturally entrenched cause-effect relations between frames. Let us begin with the commercial transaction frame, which was instantiated by the result predicates *buy* and *pay* in Table 4, and which seemed to be associated with verbs of trickery. A look at all significant collexemes of all verbs instantiating this frame confirms this preference; Table 6 shows the collexemes of commercial transaction verbs in order of association strength.

RESULT	CAUSE
buy	*mislead*, *hoodwink*, *lure*, *entice*, *boss*, *pester*, *diddle*, *guilt-trip*, *scare*, *nag*, *pressure*, *steer*, *tempt*, *fool*
purchase	*mislead*, *lure*
pay	*con*, *dupe*, *harass*, *intimidate*, *scare*, *blackmail*, *tie*, *panic*, *mislead*, *shame*
overpay	*dupe*
sell	*panic*, *force*, *entrap*, *terrify*

Table 6. Collexemes of transaction verbs in the *into*-causative

Clearly trickery verbs constitute the largest single group of cause predicates for all verbs except *sell,* to which we return presently. A second dominant group is that of coercion and coercive verbal persuasion, although the verbs from this group all encode a relatively gentle type of coercion based on continuous pressure rather than brute force (note the complete absence of verbs encoding the latter). Finally, there are a few verbs of negative emotion (*guilt-trip*, *scare*, *intimidate*, *panic,* and *shame*). Thus a cultural model emerges of the buyer as a passive participant in the commercial interaction, exploited (and relatively easily so) by others for their own gain. In contrast,

the prevalence of verbs of strong coercion and strong negative emotion for *sell* could be evidence that the seller is seen as a more self-determined agent, who is less easily coerced than the buyer.

As the commercial transaction frame is strongly associated with the TRICKERY frame, the question arises whether the reverse association also holds; even though verbs of trickery are the strongest collexemes of transaction verbs, the latter are not necessarily among the strongest collexemes of the former. A look at the trickery verbs appearing with at least two transaction verbs, however, shows that the reverse association holds (cf. Table 7). Although, of course, people can be lured, misled, and duped into a range of activities, *buy* is the top collexeme for *lure* and *mislead*, and *mislead* co-occurs nearly exclusively with commercial transaction verbs

CAUSE	RESULT
lure	*buy*, *watch*, *purchase*, *play*, *try*, *say*, *attend*, *set*, *confess*
mislead	*buy*, *purchase*, *vote*, *pay*
dupe	*carry*, *leave*, *telephone*, *pay*, *appear*, *overpay*, *hand*, *answer*, *attend*

Table 7. Collexemes of *lure*, *mislead*, and *dupe* in the *into*-causative

Next, let us take a look at the frames involved in the second-most significant combination *torture-confess*. The collexemes of *confess* are, in order of association strength: *torture*, *beat*, *intimidate*, *trap*, *coerce*, *lure*. Briefly, the CONFESSION frame in question (which is not the religious one) involves two participants; a confesser and a confessed-to, where the former admits to doing something that is judged wrong by the latter. Clearly a confession is not in the interest of the confesser, and thus it makes sense that potential causes are events of prolonged physical violence or of trickery. It seems that violence is more strongly associated with confession than trickery, however, and a look at the reverse associations confirms this. Table 8 lists the significant collexemes of each of the significant collexemes of *confess*.

CAUSE	RESULT
torture	*confess*, *betray*, *tell*
beat	*confess*
intimidate	*keep*, *leave*, *abandon*, *confess*, *vote*, *hand*, *stay*, *refuse*, *pay*, *approve*, *drop*
trap	*export*, *say*, *marry*, *feel*, *confess*
coerce	*have*, *cooperate*, *change*, *work*, *return*, *reverse*, *confess*

Table 8. Collexemes of some verbs of coercion in the *into*-causative

For *torture* and *beat*, the association is extremely strong in this direction as well: *confess* is the strongest collexeme, and in fact the only one for *beat*; furthermore, the second-ranked collexeme of *torture* is semantically similar to *confess* in involving a communicative act which is not in the interest of the speaker. For *intimidate*, the association is still relatively strong, with *confess* as the fourth-strongest collexeme. The strongest collexeme, *keep*, occurs exclusively as part of the phrasal verb *keep quiet*; there is thus an interesting contrast between *torture* (and *beat*), which involve inflicting pain on the patient with the promise to stop the pain upon compliance, and which are most strongly associated with getting a person to say something they do not want to say (cf. *confess* and *betray*); and *intimidate*, which involves fear and a loss of confidence on the part of the patient, and which is strongly associated with getting a person to not say something they *do* want to say. Finally, for the next three collexemes, the association with the CONFESSION frame is much weaker; here *confess* is the least strongly associated verb among the significant collexemes.

As a final example, let us look at two semantically related verbs that differ substantially in their covarying cause-predicate collexemes, *help* and *support* (as in *An undercover cop blackmails Alex into helping him* and *Schröder attempted to bounce Blair into supporting his nominee*). The collexemes of *help* are, in order of significance, *woo*, *pressgang*, *seduce*, and *press*; those of *support* are *hijack*, *bounce*, *frighten*, *drag*, and *galvanize*. It is interesting that there is no overlap between the collexemes of these verbs; furthermore, note that *support* takes exclusively negative stimuli suggesting a high degree of mental or physical violence, while *help* also admits positive verbal persuasion (*woo*) and trickery (*seduce*). This may be related to the fact that *help* denotes an activity that says nothing about the mental state of the helper, while *support* requires the supporter to believe in what s/he is doing; the relatively disinterested act of helping thus requires less force to be brought about than the relatively involved act of supporting, which involves, first of all, inducing a supportive mental state in the causee.

4 Conclusion

In this paper we have demonstrated an addition to our family of collostructional methods, which now consists of (i) the identification of collexemes (Stefanowitsch and Gries 2003), (ii) the identification of distinctive collexemes (Gries and Stefanowitsch, 2004), and (iii) the analysis of the covariation of collexemes in different slots of a construction. Our study of the *into*-causative illustrates that the choices speakers make when filling slots

in a construction are influenced by statistically significant interdependencies between these slots.

The regularities observed for the *into*-causative can essentially be accounted for by recourse to two different factors: a cognitive factor, involving different construals of causative events; and a cultural factor, involving culturally entrenched cause-effect relations between semantic frames. Since the *into*-causative overwhelmingly encodes a certain type of social interaction, it comes as no surprise that the cultural factor seems to be dominant for this construction. Our ongoing research on other constructions such as the *way* construction indicates, however, that sometimes a cognitive explanation—in this case an image-schematic one—is more rewarding. The analysis of more constructions with two slots will hopefully shed light on further regularities governing the dependencies between different constructional slots. Incidentally, note that the method is not restricted to the analysis of two slots, but can be extended to constructions with more than two slots or even to slots in frames instantiated by several different constructions (extending the type of analysis in, e.g., Atkins 1995).

A few potential refinements of our method are worth mentioning. One is concerned with the fact that all our data come from a single register, namely journalese. The reason for this is simply that the number of examples required for the analysis of covarying-collexemes is extremely high: even the BNC with its 100m words did not provide enough hits for the analysis of covarying collexemes; the corpus size had to be quadrupled before significant results emerged. Some constructions can therefore either be analyzed only on the basis of newspaper corpora or must await the advent of balanced giga-corpora. Once such corpora are available, a refinement that follows automatically would be to take into account the dispersion of collexemes. Given the problems inherent in basing one's analysis on collexemes that may be used rather idiosyncratically by a single speaker/writer, determining the degree of dispersion for the collexemes in question would increase the external validity of the analysis. A second refinement dependent on even larger amounts of data is the application of cluster analysis or factor-analytic techniques. These would enable the researcher to identify to what degree the type of frame-semantic knowledge posited here is reflected in multidimensional clusters of verbs. Finally, the analysis of the cultural differences manifested in covarying collexemes could be improved by investigating the differences between varieties; a case in point would be the comparison of the *into*-causative in British and American journalese. We are currently exploring these issues in more detail.

References

Atkins, B. T. S. 1995. Analyzing verbs of seeing: A frame semantics approach to corpus lexicography. *Proceedings of the twentieth annual meeting of the Berkeley Linguistics Society*, eds. S. Gahl, C. Johnson, and A. Dolbey, 42-56. Berkeley: Berkeley Linguistics Society.

Church, K. W., and P. Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics* 16: 22-29.

Church, K. W., W. Gale, P. Hanks, and D. Hindle. 1991. Using statistics in lexical analysis. *Lexical acquisition: Exploiting on-line resources to build up a lexicon*, ed. U. Zernik, 115-64. Hillsdale, NJ: Lawrence Erlbaum.

Dunning, T. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics* 19: 61-74.

Gries, S. T. 2003. Testing the sub test: An analysis of English *-ic* and *-ical* adjectives. *International Journal of Corpus Linguistics* 8: 31-61.

Gries, S. T., and A. Stefanowitsch. 2004. Extending collostructional analysis: A corpus-based approach to 'alternations'. *International Journal of Corpus Linguistics* 9:1: 97-129.

Hunston, S., and G. Francis. 2000. *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam and Philadelphia: John Benjamins.

Rohde, A. 2001. Analyzing PATH: The interplay of verbs, prepositions, and constructional semantics. Doctoral dissertation, Rice University.

Stefanowitsch, A. 2001. Constructing causation: A construction-grammar approach to analytic causatives. Doctoral dissertation, Rice University.

Stefanowitsch, A., and S. T. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8.2: 209-43.

Wierzbicka, A. 1998. The semantics of English causative constructions in a universal-typological perspective. *The new psychology of language*, ed. M. Tomasello, 113-53. Hillsdale, NJ: Lawrence Erlbaum.