Anatol Stefanowitsch

The goal bias revisited: A collostructional approach

https://doi.org/10.1515/gcla-2018-0007

Abstract: There is a goal bias in the description of motion events: adverbials specifying goals are preferred over adverbials specifying source. Two broad explanations have been suggested to account for this: first, a general cognitive bias towards the aims of human actions, and second, the higher information value of goal adverbials in conceptualizing a motion event in its entirety. The second explanation predicts that the goal bias should be verb-specific. In particular, verbs whose lexical semantics focus on trajectories or sources of motion events (such as *stroll* and *escape* respectively), should not display a goal bias but instead prefer adverbials corresponding to this focus. Stefanowitsch and Rohde (2004) present case studies of ten English motion verbs that confirm this prediction. The current study takes up this research and complements it with a collostructional analysis over a large sample of 248 English motion verbs. The study shows, first, that goal adverbials dominate among strongly-associated pairs of motion verbs and prepositions in the English Intransitive Motion Construction, confirming a general goal bias for this construction; and second, that while goal adverbials are significantly associated with generic motion verbs as well as motion verbs specifying trajectories, trajectory adverbials and goal adverbials are significantly associated with trajectory- and goal-oriented verbs respectively, adding large-scale quantitative confirmation to the previous study.

Keywords: motion events, motion verbs, goal bias, collostructional analysis, corpus linguistics

1 Introduction

There is a long-standing observation that adverbials expressing goals are preferred over adverbials expressing trajectories or sources in the linguistic description of motion events. This preference, referred to as "goal-over-source principle" (Ikegami 1979) or "goal bias" (Bourdin 1997), manifests itself in many ways. Three examples may suffice here.

Anatol Stefanowitsch, Freie Universität Berlin, anatol.stefanowitsch@fu-berlin.de

First, goals can be unmarked in the description of motion events, while sources cannot (Fillmore 1975: 26; Ikegami 1979: 143). This is a consequence of the fact that when locative prepositions are combined with motion verbs, they can be interpreted as having the semantic role of Goal, but not Source. For example, *behind the sofa* can be a description of a location, as in (1a), or of a goal (as in 1b), but must be marked explicitly with the additional preposition *from* if it describes the source (as in 1c) (first two examples from Fillmore 1975: 26):

(1) a. He left his slippers *behind* the sofa.
b. The cat ran *behind* the sofa.
c. The cat ran *from behind* the sofa to the kitchen.

Second, Source adverbials may freely be omitted from motion event descriptions, but omitting a Goal adverbial may result in semantic oddness out of context, as the following examples and judgments from Verspoor et al. (1999: 88) show:

(2) a. I climbed *from* my room *up* the ladder *onto* the roof.
b. I climbed *onto* the roof.
c. ?? I climbed *from* my room.

Third, and supporting to some extent the judgments in (2), Goal adverbials are much more frequent than Source adverbials in motion event descriptions in linguistic usage (Stefanowitsch and Rohde 2004; Lakusta and Landau 2005).

Two broad explanations have been proposed for this goal bias. First, it has been suggested that the goal bias in the encoding of motion events is a domain-specific manifestation of a more general cognitive bias towards the aims of human actions. For example, Verspoor, Dirven and Radden claim that those motion event descriptions that instantiate a "doing schema" (i.e., that have intentionally acting Themes) display a goal bias because

> a 'doing schema' by nature involves human volition and we tend to be far more interested in the goal of the action than in the source of the action. Therefore, when human action is involved, goal is far more salient than source. (Verspoor et al. 2004: 84)

Under this interpretation, the goal bias should be limited to the subclass of motion event descriptions with human (or at least animate) Themes. However, while it has been shown that a *cognitive* bias towards goals does indeed exist for such motion events, but not for ones with inanimate themes (Lakusta and Landau 2012), the goal bias is found with both types of motion events in linguistic usage (Stefanowitsch and Rohde 2004; Lakusta and Landau 2005).

Second, it has been suggested that the goal bias is a consequence of the higher information value of goal descriptions as compared to source descriptions. Ikegami (1979) phrases this suggestion in terms of expectations about the completedness of motion events:

> If we hear that something has started, we are still left with an expectation to be told that it has arrived at a certain point. Otherwise it will be felt incomplete as a description of a motion. On the other hand, if we hear that something has arrived at some place and ended its motion there, we feel quite satisfied with the description in spite of the fact that we are not told about the start of the motion. *The source and the goal are thus psychologically not equally valued as constituting elements of a completed motion.* (Ikegami 1979: 148–149; emphasis mine).

A similar explanation is found in Ungerer and Schmid (1996), who phrase it in terms of a need to infer a complete path, something that cannot be done on the basis of a Source adverbial:

> [W]hereas the medial and final portions of paths allow for an inferential conceptualization of the entire path, the information contained in the initial portions is not sufficient to establish the whole ensuing path (Ungerer and Schmid 1996: 226).

Finally, Lakusta and Landau (2012), observe that "an event that culminates in a goal presupposes some starting state (source), but not the reverse".

Under this interpretation (referred to as the "complete-conceptualization hypothesis" in Stefanowitsch and Rohde 2004), the goal bias is a linguistic rather than a cognitive bias: it assumes that we always attempt to conceptualize motion events in their entirety (with equal attention to goal, trajectory and source), and that linguistic descriptions must contain enough information to allow us to do so, which typically means that they must specify the goal.

In Stefanowitsch and Rohde (2004), we argued that the two explanations make slightly different predictions with respect to individual motion verbs. If we are generally interested in goals rather than sources of motion events, the goal bias should be observable relatively uniformly across all motion verbs. If, on the other hand, the goal bias is due to the higher information value of goal adverbials in constructing motion event descriptions, then the goal bias should be more variable. For example, it should be possible to include a Source adverbial without a Goal adverbial where the endpoint of the motion event is obvious from the context (as in 3a) or, crucially, where the lexical meaning of a motion verb focuses on the source (as in 3b) (these and all following examples are from the *British National Corpus* (BNC), the three-letter codes are the text id's):

(3) a. The travelling salesman *climbed from* his car, rolling up his sleeves as he did so. [HWN]
    b. Three prisoners *escaped from* Bois d'Arcy jail near Paris yesterday, using a helicopter. [CH2]

For our 2004 study, we chose two verbs each that seemed intuitively goal-oriented (*climb* and *flee*), two that seemed to be trajectory-oriented (*cruise* and *stroll*), and

two that seemed to be source-oriented (*fall* and *escape*). We also included four manner-of-motion verbs that we deemed, again, intuitively, to have no inherent orientation with respect to the different segments of Path (*fly*, *jump*, *run*, and *march*). We then investigated the frequency of adverbials of Goal, Trajectory and Source occurring with each of these verbs and found that, indeed, each of the verbs with a path-orientation preferred adverbials corresponding to the more specific orientation we had ascribed to them. We also found that the manner verbs differed substantially, with *fly* preferring Goal adverbials, *march* equally preferring Goal and Trajectory adverbials, *jump* preferring Goal and Source adverbials, and *run* showing no clear preference.

We took these results as evidence for the complete-conceptualization hypothesis. However, the study had a number of methodological problems that the current study will attempt to address. First, the verbs were intuitively judged to be goal-, trajectory- or source-oriented *a priori*; choosing different verbs may have led to different results. The fact that the manner-of-motion verbs *fly*, *march*, *jump* and *run* differ markedly with respect to the type of Path adverbials they prefer shows how verb-specific such biases may be. Second, the study was based on relatively small samples of each of these verbs – since they were annotated manually, we limited ourselves to 250 hits for each of the ten verbs. Finally, the study was based on a newspaper corpus, severely limiting the range of text types included.

The current paper attempts to address these issues by taking a more comprehensive and more rigorously quantified, if necessarily more superficial methodological approach: collostructional analysis as developed in Stefanowitsch and Gries (2003, 2005) and Gries and Stefanowitsch (2004). This approach captures the distribution of lexemes (in this case, motion verbs) within and across grammatical constructions (in this case, (subtypes of) the Intransitive Motion Construction). It takes into account a large set of motion verbs and postpones their assignment to semantic classes until after such associations have been determined, making it less likely to bias the result by picking – consciously or subconsciously – verbs that will support one's argument. Finally, it can be applied easily to large corpora; in this study, I will use the 100-million-word *British National Corpus* (BNC), which includes a broad range of text types.

I will present two studies, one looking at associations between individual verbs and prepositions within the Intransitive Motion Construction, and one looking at lexical associations to three subtypes of the Intransitive Motion Construction, defined by whether they contain Goal, Trajectory or Source adverbials. Before I turn to these studies, however, I will discuss in detail how the data were extracted – since there are no corpora that are annotated for argument

structure at a sufficient level of detail to extract the Intransitive Motion Construction automatically, finding a heuristic for extracting a sufficiently representative sample of occurrences from a large, non-annotated corpus is the main methodological problem to solve before a collostructional analysis of any kind can be attempted.

2 Data extraction

The Intransitive Motion Construction (henceforth: IMC) is shown in Fig. 1 based on Rohde (2001: 242), adapted from Goldberg (1995: 160). In Rohde's version, Path is intended to be a general semantic role that encompasses the more specific semantic roles Source (starting point of a path), Trajectory (segment between starting and end point) and Goal (endpoint of a path), a convention I follow throughout this study.[1]

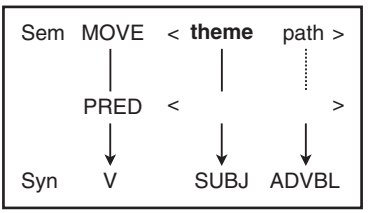

Figure 1: The Intransitive Motion Construction (IMC).

In the most straightforward case, the construction will contain a verb that lexically encodes motion, such as *move* in (4a) or *go* – an *instance* of the assumed constructional meaning. More typically, however, it contains verbs encoding the *manner* of motion, like *jump* or *run* in (4b, c). There are two broad types of Intransitive Motion constructions in the languages of the world – "path conflating" ones, that encode the path in the verb itself (for example, the Romance languages), and "manner-conflating" languages, like English and the other Germanic languages, that encode the path in an adverbial of some sort, leaving the verb slot free to add manner information (cf. Talmy 1985; Slobin 1996). Many non-motion verbs are

1 Interestingly, in Goldberg's formulation of the Intransitive Motion Construction, the semantic role of the adverbial (*oblique* in Goldberg's notation) is given as Goal, suggesting, perhaps, that this is the (proto)typical case and thus explicitly acknowledging the goal bias.

also allowed in the IMC, which will coerce a motion reading in these cases – for example, sound-emission verbs like that in (4d).

The adverbial in the IMC is typically a PP headed by directional preposition that may be transitive (as in 4a, d) or intransitive (as in 4b)[2], and may encode any element of the path – the goal (4a), the trajectory (4d) or the source (4b). However, it can also be a locative preposition, as in (4c), in which case a path reading is coerced by the construction.

(4) a. Tanks and firehose trucks *moved into* the square ... [AAL]
b. The [...] policeman attempted to speak to him but the child *ran away*. [A5Y]
c. [S]he *jumped in* the car and started it up with a violent rev of the engine. [JYC]
d. [T]he train *rumbled through* the night towards the north. [GVT]

Clearly, it is not straightforwardly possible to extract this construction from a corpus automatically with a reasonable degree of precision and recall. At the phrasal level, its instances should adhere to the structure in (5) (taking into account that the preposition may be optionally modified by adverbs, e.g. *walk **straight** into a bar*):

(5) V_{intr} (Adv) $P_{spatial}$

However, while cases of the IMC should generally adhere to the structure in (5), the converse is not true: instances of the structure in (5) will not uniquely correspond to the IMC, but will include all intransitive verbs with spatial (or temporal) prepositional complements and adjuncts. In order to maximize recall, we would have to construct a query corresponding to (5) and then manually extract the hits corresponding to the IMC. The precision of such a query is quite low: even restricting the query to the more straightforward cases with a transitive spatial preposition followed by an NP, it is only 0.32 (a sample of 100 hits from the BNC yields 32 (literal and metaphorical) cases of the IMC; see Appendix A3.1 for the CQP query used).

Since even the query restricted to transitive PPs yields 886,833 hits, manual extraction is not feasible unless precision is increased. The only way to achieve this is to limit the query to verbs encoding motion lexically. This will exclude, for

2 I follow Huddleston and Pullum (2002) in treating words like *away*, *back*, *downhill*, *out*, *upwards* etc. as intransitive prepositions, so the term *preposition* should be understood in the following to refer to what is traditionally referred to as "spatial adverbs" as well as to what is traditionally referred to as "prepositions".

example, cases like (4d), where the motion reading of the verb is coerced by the construction. This could have a relevant impact on our results: if the IMC *does* have a general preference for goal adverbials, it is possible that lexical motion verbs will differ from coerced motion verbs with respect to this preference. Specifically, Goschler (2011) shows that noise-emission verbs in German are less likely than lexical motion verbs to occur with telic adverbials. If this is also true for English, excluding coerced verbs may over-represent the goal bias in our sample somewhat.

In order to restrict the query to verbs encoding motion lexically, two existing lists of motion verbs were combined (Lawler 1999 and Levin 1993). For each verb, it was determined by a superficial inspection of a KWIC concordance based on the BNC, whether motion uses were substantially present. A number of verbs were removed because their motion uses were deemed too marginal. In addition, *arrive* was excluded because – unlike any other motion verb – it is used mainly with the preposition *at*, which was excluded here for the reasons discussed below. Including *arrive* while not including the Goal preposition *at* would distort the results, as *arrive* would falsely appear to be associated with Path or Source adverbials. This left the verbs listed in Appendix A1 (which also lists the removed verbs).

A second way in which the query in (5) has to be restricted is with respect to the prepositions included. Specifically, locative prepositions which retain their locative meaning even when used with motion verbs have to be excluded as they would distort the result substantially. For example, *in* is interpreted as directional in (4c) above, repeated here as (6a), but as locative in (6b) and, by metaphorical extension, as temporal in (6c):

(6) a. [S]he *jumped in* the car and started it up with a violent rev of the engine. (=4c)
b. A fish *jumped in* the water. [HTM]
c. We ... fitted our parachutes and decided to *jump in* the failing light. [AR8]
d. He sent for his carriage and *jumped in*. [A18]

In particular, transitively used *in*, *on* and *at* retain their locative meaning in the majority of cases even when used with motion verbs and must be removed from the query in order to increase precision to an acceptable level (intransitive uses of *in* and *on* can be included, as they always have a Goal reading, cf. 6d). However, as discussed in the introduction, whenever these prepositions are not interpreted as locative, they receive a Goal interpretation. Thus, removing them from the query will under-represent the goal bias in the resulting sample.

In order to limit the query to directional prepositions, I consulted Lindstromberg (1998) and the extensive discussion of prepositional meanings in Quirk et al. (1985: 672–687) and drew up the list of directional prepositions in

Appendix A2. Where a preposition was dominantly locative in its transitive use but directional in its intransitive use (like *in* or *on*), it was included in the query such that only intransitive uses would be captured (see further below). Note that the list contains some complex prepositions, like *out of* and *away from* – these were included based on Lindstromberg's (1998) discussion, as were multi-word expressions like *all across*. The latter differ from their simple counterparts in that they imply non-directed trajectories; for example, *We walked across the park* means that we walked from one end to the other, while *We walked all across the park* means that we walked around *in* the park. The distinction between directed and non-directed trajectories is not important in our context, but it may become so in future research.

Taking these two restrictions into account, and based on the respective lists, a query was constructed to correspond to the following structures, where (7a) must not be followed by an NP, and neither (7a) nor (7b) may be followed by another directional preposition within the same clause (as cases of the IMC with more than one PP (like example 2a above) are not relevant to the investigation here):

(7) a. V_{motion} (Adv) $P^{intr}_{directional}$ ...
b. V_{motion} (Adv) $P^{trans}_{directional}$ NP ...

The complete query is provided in Appendix A3.2 in the CQP language using the CLAWS 5 tag set. It will produce the data set used here if applied to the version of the BNC prepared for the Corpus Work Bench (Evert and Hardie 2011) using the scripts provided by Evert (2016). Roughly, it works as follows:

(8) i. a) Look for a word whose lemma is one of the words in the list of motion verbs in Appendix A1 and which is tagged as a verb; if the verb is tagged as a past participle (VVN in CLAWS 5), then it must b) be preceded by the auxiliary *have*, an optional negative particle or clitic, and zero or more adverbs;

The condition in b) is introduced in order to limit past participles to present and past perfect contexts, thus excluding passives of transitive uses of the motion verbs. For example, the query should find *Every man has moved into position* [A77], but not *Albert was moved into a nursing home* [A7C].

ii. Allow (but do not require) one of a list of adverbs to occur, that frequently appear with motion verbs but are not themselves goal- or source-oriented;

Specifically, these were the adverbs *again, almost, already, also, before, close, closer, deep, deep, deeper, ever, far, farther, fast, further, halfway, hard, headlong, instead, just, later, less, more, now, once, only, overnight, quite, right, roughshod, slow, so, solo, still, straight, today, together, too, twice, way, well, yesterday, yet*, and all adverbs ending in *–ly*. This list was created by running a version of the query that looked for all adverbs

occurring in this position and then selecting manually those that occur at least five times and do not have a directional reading.

iii. Require a transitive or intransitive directional preposition;
a) if the preposition is transitive (PRP in CLAWS 5), it must be followed by an NP
b) if the preposition is intransitive (AV0 or AVP in CLAWS 5), it must not be followed by an NP
The condition in a) is achieved by specifying a macro in CQP that captures a sequence of part-of-speech tags consisting either of a pronoun, or of the sequence (Det) (Adj) N+, or of the sequence (Det) (Adj) N 's (Adj) N+; this does not capture all, but certainly the most frequent types of NP in English. The condition in b) is achieved by specifying that the word following an intransitive preposition must not be of one of the word classes that may signal the beginning of an NP (i.e., not a pronoun, determiner, numeral, adjective, or noun).

iv. Allow any sequence of words other than a potentially directional preposition until you hit the end of the current clause.
This condition ensures that instances of the IMC that contain more than one Path PP are excluded. The end of the current clause is assumed to be reached either if the query hits an instance of sentence-final punctuation or a verb. Removing this condition will allow the query to also find cases of the IMC with more than one PP

The query in (8) is obviously still not perfect and may be further improved by checking the results for systematic problems (see the remarks in Study 1), but its precision is high: in a sample of 100 hits (see Appendix A3.2), there were only three hits that do not exemplify the IMC: *run out of willpower*, *leave aside the question of …*, and *come up with something*. The precision in the smaller sample is thus a very satisfactory 97 percent, and is likely to be over 90 percent in the larger sample, allowing us to forego manual extraction and treat the false hits as noise. However, it should be noted that the query captures both literal and metaphorical uses of the motion verbs in question, since there is evidence that the latter do not display a goal bias to the same extent as the former (Stefanowitsch and Rohde 2004), this may lead to an under-representation of the goal bias in the data set derived by this query. Since the goal bias itself is not at issue here, this should not be a major problem, however. Of course, metaphorical uses of motion verbs may have other properties that influence the results presented here; but note that even distinctly phraseological uses of motion verbs outside of the IMC can be systematically related to their literal meaning (cf., e.g., Schönefeld 2013).

The query yielded 185,923 hits. These were then post-processed to yield a frequency list of combinations of verbs and prepositions which served as the basis for both of the following studies (the full frequency list is available in the Supplementary Materials).

3 Corpus study 1: Covarying motion verbs and prepositions in the IMC

3.1 Aims and method

The aim of this study is to identify strongly associated pairs of motion verbs and prepositions in our sample in order to determine, first, whether there are such pairs at all and whether goal-oriented prepositions dominate among the most strongly associated pairs (as the existence of a general goal bias would lead us to expect), and second, whether exceptions (specifically, pairs with source-oriented prepositions) are also found among the most strongly associated pairs and whether these can be explained with reference to the lexical semantics of the verb (as the complete-conceptualization hypothesis suggests).

The method used is a covarying-collexeme analysis (Stefanowitsch and Gries 2005) between the motion verbs and the prepositions. The analysis was performed using the R package {collostructions} (Flach 2017), with the G^2 measure of the log-likelihood test as a test statistic.

3.2 Results

There are 1,077 significantly attracted verb-preposition pairs (i.e., pairs that are significantly more frequent than expected) and 604 significantly repelled pairs (i.e., pairs that are significantly less frequent than expected). In addition, there are 2,079 pairs whose association is not significant in the sample.[3]

Clearly (and not unexpectedly), motion verbs have preferences for particular prepositions within the IMC. The top twenty-five covarying verbal and prepositional collexemes are shown in Table 1.

Some of these pairs are false hits, showing that the query used for this paper is still not perfect. For example, eighth-ranked *speed up* is a particle verb (with both transitive and intransitive uses), where *up* does not have a directional meaning; there are very rare cases where the sequence *speed up* represents the IMC (e.g. *By nine a.m., Detective Inspector Richard Montgomery [...] and his trusty assistant [...] were speeding up the same road* [C8D]), but overwhelmingly, it represents the phrasal verb. Similarly, the sequence *wind up* represents the IMC in only about 10 percent of all cases, with the remainder representing the phrasal verb *wind up* ("arrive at a

3 The complete data set is available in the Supplementary Materials.

Table 1: Significantly covarying verbs and prepositions in the IMC.

Verb	Preposition	f(Verb)	f(Prep)	O(Comb)	E(Comb)	G^2
go	*on*	69,078	17,216	11,431	6,396.4	6,680.56
come	*from*	56,939	9,454	6,381	2,895.3	5,753.64
go	*to*	69,078	39,247	20,163	14,581.9	4,200.29
enter	*into*	1,111	16,142	989	96.5	4,142.44
escape	*from*	784	9,454	567	39.9	2,507.43
come	*back*	56,939	8,490	4,697	2,600.1	2,334.49
go	*ahead*	69,078	1,790	1,498	665.1	1,661.29
speed	*up*	605	10,900	402	35.5	1,546.39
come	*in*	56,939	7,906	3,928	2,421.2	1,306.10
leave	*home*	1,108	4,581	341	27.3	1,218.74
move	*towards*	8,346	2,344	572	105.2	1,136.41
leave	*aside*	1,108	292	148	1.7	1,133.69
run	*away*	7,109	5,076	761	194.1	1,063.91
wind	*up*	409	10,900	274	24	1,058.10
depart	*from*	239	9,454	204	12.2	1,024.09
move	*away from*	8,346	1,167	370	52.4	924.85
swing	*round*	490	2,556	164	6.7	800.46
roll	*over*	947	2,242	174	11.4	665.04
move	*between*	8,346	691	228	31	586.28
struggle	*against*	350	691	84	1.3	566.55
run	*out*	7,109	15,623	1,195	597.4	544.77
travel	*to*	1,405	39,247	690	296.6	543.34
hurry	*up*	860	10,900	269	50.4	533.53
step	*forward*	2,347	1,910	201	24.1	529.85
walk	*towards*	6,168	2,344	338	77.8	515.59

place/situation"). Such combinations should probably be removed altogether from an improved query. In other cases, the uses corresponding to the IMC are more frequent, if still in the minority. For example, the sequence *run out* represents the IMC (literally or metaphorically) about a third of the time, with the phrasal verb *run out* ("be used up") accounting for the rest. Finally, there are cases where the preposition is directional but forms part of a transitive phrasal verb. For example *leave aside* represents uses like *Leaving aside the areas of conflict, librarians would like to believe that their input to the trade is a vital one* [B07], which is a case of the Caused-Motion Construction with *areas of conflict* as a theme in object position rather than subject position. None of these cases have very high token frequencies, however, so they should not have an undue impact on the overall result; even the relatively frequent sequence *run out* accounts for only 0.6 percent of the data (1,195/185,923). The majority of the top thirty covarying collexemes in Table 1 (and in the data set as a whole) represent the IMC, confirming the general usefulness of the query used here.

Goal-oriented prepositions are in the majority among the top co-varying verb-preposition collexemes that represent the IMC at least a third of the time: *go on*, *go to*, *enter into*,[4] *come back*, *go ahead*, *come in*, *struggle against*, *travel to*, *step forward*, and *walk towards*. However, source-oriented prepositions are also substantially present: second-ranked *come from*, as well as *escape from*, *run away*, *depart from*, *move away from*, and *run out*. Trajectory prepositions are in the minority (*swing round*, *roll over*, *move between*).

3.3 Discussion

The results of this study can be broadly interpreted as evidence that the IMC in English is subject to the goal bias. They leave open the question whether exceptions to this are to be found mainly in the semantics of individual verbs, as the complete-conceptualization hypothesis predicts. Among the source-oriented verb-preposition pairs in Table 1, two have a clearly source-oriented verb (*escape from* and *depart from*), the other pairs have verbs encoding generic motion (*move away from*), deictic trajectories (*come*), or manner of motion (*run*). Among the (more frequent) goal-oriented pairs, only a single verb has a clear goal orientation (*enter*). This suggests that exceptions to the goal bias may be found more frequently with source-oriented verbs, but only tentatively. In order to investigate this possibility more systematically, we have to move beyond individual verb-preposition pairs and compare the verbs associated with goal-, trajectory- and source-prepositions in general. Study 2 is an attempt to do this.

4 Corpus study 2: A distinctive-collexeme analysis of Goal, Trajectory and Source in the IMC

4.1 Aims and method

The aim of this study is to identify motion verbs from the sample used here that are significantly associated with one (or two, see below) of three variants of the

4 Note that *enter into* almost exclusively represents metaphorical uses of the verb *enter*: in Modern English, literal uses of *enter* overwhelmingly encode the Entered Location as a direct object, while metaphorical uses have retained the Goal adverbials initially found with literal uses also (cf. Stefanowitsch 2013).

IMC: a) the IMC_{GOAL}, where the adverbial has the semantic role Goal (*They climbed into the back of the car* [A6N]), b) the IMC_{TRJCT}, where the adverbial has the semantic role Trajectory (*Willie climbed up the ladder* [CAB]), and c) the IMC_{SRC}, where the adverbial has the semantic role Source (*The travelling salesman climbed from his car*, cf. (3a) above). I do not assume here that these are necessarily different subtypes of the IMC in the sense that they have properties not predictable from the IMC itself and the type of adverbial (although I am open to this possibility). Positing these three subtypes is merely an analytic device to investigate the role of Source, Trajectory and Goal adverbials with respect to whether these classes of adverbials show systematic associations with particular verbs and, crucially, verb classes. The specific research question here is whether significant verbal collexemes of the IMC_{TRJCT} and the IMC_{SRC} are more likely to encode lexically, respectively, a trajectory- and source-orientation, while the verbs associated with the IMC_{GOAL} are more likely to be generic motion verbs or manner verbs that do not entail or imply a goal-orientation.

In order to investigate lexical associations within a delimited set of constructions, some variant of distinctive collexeme analysis (Gries and Stefanowitsch 2004) is typically used. In the simplest case of a comparison between two constructions, this is straightforward: a two-by-two contingency table is constructed for each verb, where its frequency in each of the two constructions is compared to the frequency of all other verbs in each of the two constructions. When there are more than two constructions (say, three, as in our case), there are two possible ways to extend this analysis. Either we increase the dimensionality of the table and compare the frequency of each verb in each of the three constructions against the frequency of all other verbs in each of the three constructions and identify the cell in which the difference between the observed and expected frequencies is largest. Or we conduct three separate analyses for each verb, comparing its frequency in each of the three constructions against its combined frequency in the other two – this essentially extends the logic of the verb slot, where we always compare the frequency of one particular verb against the combined frequency of all other verbs, to the construction itself.

For the purpose of this study, I chose the latter strategy, conducting the analysis by recoding all prepositions as either Goal, Trajectory or Source as shown in Appendix A2 and then running a covarying collexeme analysis over verbs and recoded prepositions using, again, the R package {collostructions} (Flach 2017). This method seems both simpler and more in line with the conceptual design of the current study, which essentially investigates the co-variation between two slots in a construction, even if one of the slots enters the analysis in the form of semantic roles rather than individual lexical items.

4.2 Results

There are 125,129 tokens of the IMC_{GOAL} in the data, distributed across 198 types (i.e., 198 different motion verbs from our sample). This makes it the most frequent of the three variants by far – again, a broad confirmation of a general goal bias. There are 26,374 tokens of the IMC_{TRJCT} in the data, distributed across 191 types, making it the least frequent one of the three versions of the IMC. Finally, there are 34,420 tokens of the IMC_{SRC} in the data, distributed across 171 types. This version of the construction is more frequent than the IMC_{TRJCT}, but still much less frequent than the IMC_{GOAL}. Let us look at each of the three subtypes in turn, beginning with the IMC_{GOAL}.[5]

Table 2 lists all significant collexemes of the IMC_{GOAL} – there are only 19,[6] although *arrive* would also be on this list if we had not excluded it.

Table 2: Significant collexemes of the IMC_{GOAL}.

Word	O(IMC)	O(IMC_{GOAL})	E(IMC_{GOAL})	G^2
go	69,078	52,647	46,490.5	4,074.62
enter	1,111	1,043	747.7	468.53
wind	409	390	275.3	197.93
plunge	246	227	165.6	88.54
step	2,347	1,784	1,580	86.48
speed	605	471	407.2	32.87
creep	633	479	426.0	21.29
struggle	350	267	235.6	13.61
hurry	860	623	578.8	10.69
shin	13	13	8.7	10.30
come	56,939	38,599	38,320.8	8.92
stump	44	38	30	8.46
journey	70	56	47	5.60
ski	33	28	22.2	5.28
descend	265	195	178.3	4.94
lurch	138	104	92.9	4.29
move	8,346	5,702	5,617.0	4.14
somersault	12	11	8.1	4.06
sashay	5	5	3.4	3.96

5 The complete result of the analysis is available in the Supplementary Materials.

6 Since the IMC_{GOAL} is the most frequent construction in the sample, it will have the fewest significant collexemes, as larger deviations from the expected frequencies are necessary for an association to reach significance (Stefanowitsch and Gries 2003).

There are three general motion verbs among the top twenty collexemes of the IMC_{GOAL} (i.e., verbs that encode nothing about the motion event itself other than spatial displacement): first-ranked *go*, as well as *move*, and *journey* – note that the latter suggests certain intentions behind the motion event, such as leisure or exploration, but does not entail anything about the motion event itself, e.g. in terms of speed, manner, directionality etc. The association of these verbs is unique to the IMC_{GOAL}, they are negatively associated with the other two variants of the IMC.

There is one verb in the data that lexically encodes the goal-orientation present in the semantics of the IMC_{GOAL}, *enter*, although *arrive*, excluded in the analysis, must also be included here.

Four verbs encode some aspect of Trajectory: *come* does so very generally, in that it is used deictically with the speaker as the reference point; *plunge* and *descend* encode situations where the Theme ends up in a lower position than the one in which it started, and thus entail a downward direction; finally, *wind* encodes a trajectory that changes direction frequently and randomly.

The remaining eleven verbs fall into the category manner-of-motion verbs that are typologically expected in English: six encode particular motor patterns (*step*, *shin*, *stump*, *ski*, *somersault*, *sashay*), four encode aspects of the speed of motion (*speed*, *creep*, *hurry*, *lurch*), and one, *struggle*, encodes the fact that the motion is impeded by external obstacles or an internal lack of energy.

Next let us turn to the IMC_{TRJCT}. This variant has the highest number of significantly attracted collexemes, 116, the top 20 of which are shown in Table 3.

Table 3: Significant collexemes of the IMC_{TRJCT}.

Word	O(IMC)	$O(IMC_{TRJCT})$	$E(IMC_{TRJCT})$	G^2
wander	620	365	87.9	667.21
walk	6,168	1,617	875.0	636.72
swing	490	252	69.5	379.70
drive	1,871	563	265.4	314.46
run	7,109	1,537	1,008.4	298.43
whirl	107	78	15.2	188.67
travel	1,405	363	199.3	132.39
roam	58	46	8.2	124.27
flit	61	47	8.7	122.21
fly	1,689	408	239.6	119.20
roll	947	244	134.3	87.97
float	181	77	25.7	85.82
rush	1,018	253	144.4	81.17
prowl	27	24	3.8	75.84
skip	95	48	13.5	70.23
twist	111	52	15.7	67.78

Table 3 (continued)

Word	O(IMC)	O(IMC_{TRJCT})	E(IMC_{TRJCT})	G^2
ride	608	163	86.2	66.15
shuffle	173	67	24.5	63.25
race	400	118	56.7	62.09
parade	25	21	3.5	61.28

The distribution of these verbs across the categories introduced above is clearly different from the one in the IMC_{GOAL}. Most noticeably, there is only one generic motion verb, *travel* – which, like *journey*, may encode something about the intent behind the motion event when the Theme is animate, but which does not specify anything about the motion itself, other than that it takes place.

While the IMC_{GOAL} only had two strongly associated collexemes encoding the goal-orientation inherent in it, the IMC_{TRJCT} has seven verbs that lexically encode some aspect of Trajectory – typically focusing on the fact that the motion is non-directed and/or that it is an end in itself (*wander*, *swing*, *roam*, *prowl*, *float*, *whirl*, *parade*).

The remaining twelve verbs are the kind of manner-of-motion verbs expected in the English IMC typologically – either encoding a motor pattern or other aspect of the theme directly involved in causing the motion (*walk*, *drive*, *run*, *fly*, *float*, *roll*, *skip*, *shuffle*, *ride*) or focusing on the speed of the motion event (*flit*, *rush*, *race*). In some of these cases, both motor pattern and speed are encoded (*walk* vs. *run*, *shuffle* vs. *fly*, etc.).

Finally, let us look at the IMC_{SRC}. This variant has 41 significantly attracted collexemes, the first twenty of which are shown in Table 4.

Table 4: Significant collexemes of the IMC_{SRC}.

Word	O(IMC)	O(IMC_{SRC})	E(IMC_{SRC})	G^2
run	7,109	2,781	1,316.1	1,718.18
escape	784	573	145.1	1,112.73
depart	239	205	44.2	510.87
slip	1,511	657	279.7	502.19
come	56,939	12,172	10,541.1	437.42
stretch	657	291	121.6	230.33
drop	1,133	389	209.8	160.49
exit	38	33	7.0	83.80
storm	189	86	35.0	71.90
wriggle	116	62	22.0	71.06

Table 4 (continued)

Word	O(IMC)	O(IMC_{SRC})	E(IMC_{SRC})	G^2
spring	686	198	127.0	43.46
flee	444	140	82.2	43.39
walk	6,168	1,342	1,141.9	42.74
leap	824	228	152.5	41.37
stray	176	63	33.0	29.24
sweep	379	111	70.2	25.87
jump	1,417	336	262.3	23.98
tumble	236	75	43.7	23.87
climb	1,206	289	223.3	22.34
slog	16	11	3.0	19.28

The distribution of these verbs across the categories introduced above is more similar to those in the IMC_{TRJCT} than the IMC_{GOAL}. Most strikingly, there are no generic motion verbs at all among the top twenty (or indeed, among any of the) significant collexemes of the IMC_{SRC}.

Instead, there are five verbs that lexically encode the source-orientation inherent in the IMC_{SRC} (*escape*, *depart*, *exit*, *flee*, *stray*), and four verbs encoding some aspect of Trajectory: the deictic *come* also associated with IMC_{GOAL}, and three verbs encoding trajectories where source and goal are typically at different heights and which thus imply motion along a vertical dimension (*drop*, *climb*, *tumble*).

Most of the remaining verbs are the manner-of-motion verbs typical of the IMC in English, some of them encoding motor patterns (*walk*, *wriggle*, *leap*, *slog*, *run*, *jump, stretch*), and some focusing on some aspect of speed (*slip*, *storm*, *spring*, *sweep*). Again, some verbs combine speed and motor pattern, such as *walk* vs. *run*.

4.3 Discussion

Before we summarize the differences between the three variants of the IMC, note that they are broadly similar in one respect: for all three of them, slightly more than half of the significantly attracted collxemes consist of manner-of-motion verbs. This is expected, since manner-of-motion verbs are characteristic of manner-conflating languages like English. Still, it is interesting to note that, in a direct comparison of the three variants, each one attracts its own, relatively unique set of motion verbs, a point I will briefly return to in the Conclusion.

Let us turn to the verbs encoding some aspect of the goal, trajectory or source of a motion event which are the focus of this study. With respect to these verbs, there is a broad difference in that such verbs are more frequent overall for the IMC_{GOAL} (which has 9 such collexemes among the top twenty) than for the other two variants (which have 6 and 7 respectively). Crucially, there is a clear difference with respect to the specific segment of the path that the verbs encode: all three variants of the IMC have trajectory-oriented verbs among their top collexemes, but while the IMC_{SRC} has five source-oriented verbs among its top twenty collexemes, the IMC_{GOAL} only has two goal-oriented ones. This suggests that exceptions to the goal bias may indeed be due to the semantics of specific motion verbs: if the verb is source-oriented, it is more likely that only the source of the motion is encoded in the IMC; likewise, if the verb is trajectory-oriented, especially if the trajectory is one without clear directionality, it is more likely that only the trajectory will be encoded in the IMC. Where the goal is encoded in the IMC, the semantics of the verb do not matter, as demonstrated most clearly by the fact that both of the generic motion verbs in English, *go* and *move*, are significantly (and exclusively) attracted to the IMC_{GOAL}. Put differently, the IMC_{GOAL} can be considered the default variant of the IMC.

5 Conclusion

This paper has added further evidence to the well-established fact that there is a goal bias in the encoding of motion events, specifically, motion events encoded by the English Intransitive-Motion Construction. This goal bias is evident if we investigate specific verb-preposition pairs in the IMC, where a clear majority of the strongly associated pairs include a goal-oriented preposition. It is also evident if we investigate associations between motion verbs and three variants of the IMC, one with goal adverbials, one with trajectory adverbials and one with source adverbials. Such an investigation shows, first, that Goal adverbials in the IMC have no strong preference for goal-oriented verbs, while generic motion verbs are strongly associated with Goal adverbials, suggesting that Goal adverbials are the default in the IMC; and second, that Trajectory adverbials have a noticeable preference for trajectory-oriented verbs and Source adverbials for source-oriented verbs, suggesting that deviations from the goal-orientation of the IMC occur in situations where the focus is uniquely on the trajectory or the source – situations that are more likely to be encoded by verbs with a corresponding focus.

Thus, the studies presented in this paper supports the conclusions drawn from the much more carefully detailed, but also much narrower analysis in

Stefanowitsch and Rohde (2004), i.e. they support the idea that the goal bias is a consequence of the fact that Goal adverbials normally have a higher information value than Trajectory or Source adverbials and are preferred for this reason, but that Trajectory and Source adverbials may have a higher information value either contextually, or, crucially, with specific verbs which will therefore be biased towards these rather than towards goal descriptions.

The analysis also showed, interestingly but not surprisingly, that manner-of-motion verbs make up the majority of significantly-attracted collexemes for all three variants of the IMC. Future, more detailed analyses of the specific type of manner verbs associated with the three variants might well uncover interesting differences between them or shed further light on the semantics of the verbs involved. Why, for example, is *journey* associated with the IMC_{GOAL} while the roughly synonymous *travel* is associated with the IMC_{TRJCT}? Or, is there a difference between the manner-of-motion verbs associated with the different variants in terms of the type of manner or the richness of the manner component encoded by the verb? These and other questions await further investigation.

Finally, I hope that this paper also contributes methodologically to the study of motion events. Motion constructions, particularly the IMC, are among the constructions which have not received much attention in the collostructional literature, despite the fact that they constitute a central area of research in cognitive and usage-based linguistics. This is no doubt due to the fact that these constructions are difficult to retrieve from non-annotated corpora. The query crafted for this paper, while it can and must still be improved, should provide a step towards solving this problem, which will enable researchers to investigate the questions raised above as well as many other questions about the IMC quantitatively and based on large corpora.

Data

The British National Corpus, version 3 (BNC XML Edition). 2007. Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. Online: www.natcorp.ox.ac.uk/

Supplementary material

Supplementary materials including data and scripts are available under a Creative-Commons-Attribution-International 4.0 license via the Open Science Framework at https://osf.io/mpa6b/.

References

Bourdin, Philippe. 1997. On goal-bias across languages: Modal, configurational and orientational parameters. In Bohumil Palek (ed.), *Typology: Prototypes, item orderings and universals* (Proceedings of LP '96), 185–218. Praha: Charles University Press.

Evert, Stefan & Hardie, Andrew. 2011. Twenty-first century Corpus Workbench: Updating a query architecture for the new millennium. In *Proceedings of the Corpus Linguistics 2011 Conference*, 1–21. Birmingham: University of Birmingham.

Evert, Stefan. 2016. *The IMS Open Corpus Workbench (CWB): CQP query language tutorial — CWB version 3.4*. Online: cwb.sourceforge.net

Fillmore, Charles J. 1975. *Santa Cruz lectures on deixis 1971*. Bloomington: Indiana University Linguistics Club.

Flach, Susanne. 2017. {collostructions}. An R implementation for the family of collostructional methods. R package version 0.1.0. http://sfla.ch.

Goldberg, Adele. 1995. *Constructions: A Construction Grammar approach to argument structure*. Chicago & London: The University of Chicago Press.

Goschler, Juliana. 2011. Geräuschverben mit direktionaler Erweiterung: Syntax, Semantik und Gebrauch. In Alexander Lasch & Alexander Ziem (eds.), *Konstruktionsgrammatik III: Vom Forschungsparadigma zu Fallstudien*, 29–43. Tübingen: Stauffenburg.

Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9(1). 97–129.

Huddleston, Rodney D. & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge & New York: Cambridge University Press.

Ikegami, Yoshihiko. 1979. 'Goal' over 'source': A case of linguistic dissymmetry. *Hungarian Studies in English* 12. 139–157.

Lakusta, Laura & Barbara Landau. 2005. Starting at the end: The importance of goals in spatial language. *Cognition* 96. 1–33.

Lakusta, Laura & Barbara Landau. 2012. Language and memory for motion Events: Origins of the asymmetry between source and goal paths. *Cognitive Science* 36(3). 517–544.

Lawler, John. 1999. *Verbs of unaided human motion*. Online at http://www-personal.umich.edu/~jlawler/words/ [Last accessed: August 2017].

Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: The University of Chicago Press.

Lindstromberg, Seth. 1998. *English prepositions explained*. Amsterdam & Philadelphia: John Benjamins.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. Harlow: Longman.

Rohde, Ada. 2001. *Analyzing path: The interplay of verbs, prepositions and constructional semantics*. PhD thesis, Department of Linguistics, Rice University, Houston, TX.

Schönefeld, Doris. 2013. *Go mad – come true – run dry*: Metaphorical motion, semantic preference(s) and deixis. *Yearbook of the German Cognitive Linguistics Association* 1. 215–236.

Slobin, Dan I. 1996. Two ways to travel: Verbs of motion in English and Spanish. In Masayoshi Shibatani & Sandra A. Thompson (eds.), *Grammatical constructions: Their form and meaning*, 195–219. Oxford: Clarendon Press.

Stefanowitsch, Anatol. 2013. Variation and change in English path verbs and constructions: Usage patterns and conceptual structure. In Juliana Goschler & Anatol Stefanowitsch (eds.), *Variation and change in the encoding of motion events*, 223–244. Amsterdam: John Benjamins.

Stefanowitsch, Anatol & Ada Rohde. 2004. The goal bias in the encoding of motion events. In Günter Radden & Klaus-Uwe Panther (eds.), *Studies in linguistic motivation*, 249–268. Berlin & New York: Mouton de Gruyter.

Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243.

Stefanowitsch, Anatol & Stefan Th. Gries. 2005. Covarying collexemes. *Corpus Linguistics and Linguistic Theory* 1(1). 1–43.

Talmy, Leonard. 1985. Lexicalization patterns: Semantic structure in lexical forms. In Timothy Shopen (ed.), *Language typology and lexical description, Volume 3: Grammatical Categories and the Lexicon*, 57–149. Cambridge: Cambridge University Press.

Ungerer, Friedrich & Hans-Jörg Schmid. 1996. *An introduction to Cognitive Linguistics*. London: Longman.

Verspoor, Marjolijn, Rene Dirven & Günter Radden. 2004. Putting concepts together: Syntax. In Rene Dirven & Marjolijn Verspoor (eds.), *Cognitive exploration of language and linguistics* (2nd edition), 75–100. Amsterdam & Philadelphia: John Benjamins.

Appendix

A1. Motion Verbs

amble, ascend, backpack, balloon, barge, bicycle, bike, boat, bob, bobsled, bolt, boogie, bop, bounce, bound, bowl, buck, bus, cab, cancan, canoe, canter, caravan, carom, cart, cascade, cavort, charge, chariot, clamber, climb, clump, coast, come, conga, crawl, creep, cruise, crush, cycle, dance, dart, dash, depart, descend, dodder, dogsled, drift, drive, drop, enter, escape, exit, ferry, file, flap, flee, flit, float, fly, gallop, gambol, gimp, glide, go, gondola, goosestep, gyrate, hasten, helicopter, hike, hobble, hop, hurdle, hurry, hurtle, inch, jeep, jet, jig, jitterbug, jog, journey, jump, kayak, leap, leapfrog, leave, limp, lolligag, lollop, lope, lumber, lurch, march, meander, mince, moped, mosey, motor, motorbike, motorcycle, move, nip, oar, pad, paddle, parachute, parade, pedal, perambulate, pirouette, plod, plunge, polka, pounce, prance, promenade, prowl, punt, quickstep, race, ramble, rickshaw, ride, roam, rocket, roll, romp, rove, row, rumba, run, rush, sail, samba, sashay, saunter, scamper, schuss, scoot, scorch, scram, scramble, scud, scuff, scurry, scutter, scuttle, shamble, shin, shlep, shove, shuffle, shuttle, sidle, skate, skateboard, skedaddle, ski, skip, skitter, skulk, sled, sledge, sleepwalk, sleigh, slide, slink, slip, slither, slog, slop, slouch, snake, sneak, somersault, spank, speed, spring, sprint, squaredance, stagger, stalk, steam, steer,

step, stomp, storm, straggle, stray, streak, stream, stretch, stride, stroll, struggle, strut, stumble, stump, surge, swagger, sway, sweep, swim, swing, tango, tap-dance, taxi, teeter, throng, thrust, tiptoe, toboggan, toddle, totter, traipse, tram, tramp, travel, trek, trolley, troop, trot, truck, trudge, trundle, tumble, twist, vault, waddle, wade, waggle, walk, waltz, wander, weave, wheel, whirl, whiz, wiggle, wind, wobble, wriggle, yacht, zigzag, zip, zoom

Excluded:
arrive, blow, break, book, burn, chicken, clog, coach, cut, fidget, foxtrot, frolic, jive, kick, raft, rock, sag, scrape, shoot, squirm, squirrel, stamp, steal, tack, tear, toil, tootle, turn, twitch

A2. Prepositions

Goal:
aboard, abroad, aft, against, aground, ahead, ashore, aside, back, beyond, down, downhill, downstairs, forth, home, in, indoors, inside, into, on, onto, outside, sideways, to, under, up, uphill, upon, upstairs, upstream

Trajectory:
about, above, across, all across, all around, all cross, all over, all round, along, alongside, around, between, by, cross, off, over, overboard, past, round, through, via

Source:
away, away from, from, out, out of

A3. Queries

A3.1 Query corresponding to the structure in (5) in Section 2

```
# define spatial prepositions:

define $spatprep = “aboard about above a?cross against along(side)? amid(st)? among(st)? a?round at atop behind below beneath beside between beyond down from in(side|to)? near off on(to)? opposite out(side)? over past through(out)? to(wards?)? under(neath)? up upon via vis-[aá]-vis within”

# search for (active) verbs, followed by an optional adverb, followed by a transitive spatial prepostion
```

```
A = ( [pos="VH."] [pos="XX0"]? [pos="AV."]* [pos="VVN"] | [pos="VV[IDGBZ]"] )
[pos="AV."]? @[hw=RE($spatprep) & pos="(PRP)"]

# draw a random sample of 100 hits to determine the precision of the
query

randomize 66

reduce A to 100
```

A3.2 Query corresponding to the structure in (7) in Section 2

```
# define lists of motion verbs, ambitransitive, transitive and intran-
sitive directional prepositions

define $motverb = [list of verbs from Appendix A1]

define $dirprep = "aboard about a?cross along alongside a?round by
between beyond down off onto over past through under up"

define $dirpreptr = "against from into via to(wards?)? upon"

define $dirprepint = "above abroad aft aground ashore ahead aside away
back downhill downstairs home in(side)? indoors forth on out(side)?
overboard uphill upstairs upstream sideways .*wards?"

# define a noun-phrase schema

define macro np(0) '([pos="PNP"] | [pos="(AT0|DT.|DPS)" | hw="one"]*
([pos="AJ."]* [pos="N.."] [pos="POS"])? [pos="AJ."]* [pos="NP."]*
[pos="N.."]+)'

# define a query to correspond to (7a, b) in Section 2, as described
in (8)
# note: complex prepostions are queried explicity, targets are set for
the preposition part of the query

IMC = ( [pos="VH."] [pos="XX0"]? [pos="AV."]* [hw=$motverb & pos="VVN"]
| [hw=$motverb & pos="VV[IDGBZ]"] ) [pos="AV." & hw="(straight|
```

```
right|too|so|well|again|together|just|closer?|further/far|further|
far|now|only|almost|instead|deep(er)?|fast|headlong|more|twice|later|
still|roughshod|yesterday|deep|half-?way|today|ever|already|farther|
overnight|once|hard|before|also|solo|quite|slow|way|yet|less|.*ly)"]?
( ( @[hw=RE($dirpreptr) & (pos="PRP") | hw=RE($dirprep) & (pos="PRP")]
| [hw="all"] @[hw="(a?cross|a?round|over)" & pos="PRP"] | [hw="away"]
@[hw="from" & pos="PRP"] | [hw="out" & pos="(AVP|PRP)"] @[hw="of"]
) ( /np() ([pos="PRF"] /np())? ) | ( @[hw=RE($dirprepint) &
(hw!="afterwards") & (pos="AV.") | hw=RE($dirprep) & (pos="AV.")] |
[hw="all"] @[hw="(a?cross|a?round|over)" & pos="AV."] [pos!="(PNP|AT0|
DT.|DPS|AJ.|N[PN][012]|CRD)"] ) ) [hw!=RE($dirpreptr) &
hw!=RE($dirprep) & hw!=RE($dirprepint) & (hw!="[.;?!:…]") &
(pos!=".*(V[VHBDM][IDNBGZ0]|TO0|CJS|CJT).*")]* [(hw="[.;?!:…]") |
(pos=".*(V[VHBDM][IDNBGZ0]|TO0|CJS|CJT).*")]

# set the motion-verbs as keyword

set IMC keyword nearest [hw=$motverb] within left 10 words from target;

# produce a frequency list for the keyword-target combinations
# note: this must still be post-processed before it is ready for use;
adverbs must be removed and complex prepositions identified

count IMC by hw on keyword .. target

# create a reduced sample of 100 hits to check the precision of the
query

IMCred = IMC

randomize 29

reduce IMCred to 100
```