

Research article

Going to space

A cross-linguistic study using electronic corpora

Kajsa Törmä

Motion between earth and outer space can be construed in different ways in different languages. The aim of this study is to identify all prepositional constructions used to describe motion from earth to space in English and Swedish and investigate how they contribute to different conceptualizations of earth, space, and the journey in between. It also aims to highlight some challenges of using corpora for cross-linguistic research. The data was collected from Corpus of Contemporary American English and Korp and analyzed through the lenses of image-schema and construal. The study shows that there are many ways to construe this type of motion in both languages. Different parts of the journey can be put into focus, and earth and space can be construed in different ways, depending on the perspective of the speaker. The most common constructions in both languages focus on the goal and construe earth and/or space as containers.

Keywords: construal; corpus; cross-linguistic research; image schema; prepositions

Rörelse mellan Jorden och rymden kan konstrueras på olika sätt i olika språk. Syftet med denna studie är att identifiera alla prepositionskonstruktioner som används för att beskriva rörelse mellan jorden och rymden på Svenska och Engelska samt att undersöka hur de bidrar till olika konceptualiseringar av jorden, rymden och resan däremellan. Studien syftar också till att lyfta fram några utmaningar med att använda korpus för jämförande lingvistisk forskning. Den insamlade datan kommer från Corpus of Contemporary American English och Korp och analyserades med utgångspunkt i föreställningsscheman och construal. Studien visar att det finns många sätt att konstruera den här rörelsen i de båda språken. Talarens perspektiv kan försätta olika delar av resan i fokus och framställa både jorden och rymden på olika sätt. De vanligaste prepositionskonstruktionerna i bägge språk fokuserar främst på målet och framställer jorden och/eller rymden som behållare.

Nyckelord: korpus; föreställningsschema; prepositioner; komparativ lingvistik; formning

Space travel has only become a physical reality in the last century, which means that the language we use to talk about it is still in its adolescence. Motion between earth and space can be construed in several different ways, and different linguistic expressions reflect different conceptualizations of the movement. Prepositions provide key insights into these differences, since they provide information about both the destination and direction of the movement at hand. Furthermore, different languages (even closely related ones) can favor different perspectives. For example, the most common preposition in English is *into* (space), whereas the most common in Swedish is the two-word

construction *ut i* (rymden) (lit. “out in space”). In English, the focus is firmly on the destination (space) which is construed as a container one can enter. Conversely, in the Swedish expression, the point of departure (earth) is highlighted and construed as a container one can leave. Differences such as these are intriguing and warrant a more in-depth examination. However, English is more widely studied than Swedish, and the available resources for studying English are far more exhaustive. Consequently, the aim of this study is twofold. Firstly, it aims to investigate prepositional constructions used in descriptions of motion from earth to space in English and Swedish. Secondly, it aims to

highlight some challenges involved in the research process. The first part of the aim will be addressed via the following research questions:

1. Which prepositional constructions are most common in English and Swedish when expressing motion to *spacelrymden*?
2. What different conceptualizations of earth, space, and the journey in between do these prepositional constructions reflect?

The secondary aim will be achieved by chronicling the research process in careful detail, outlining the limitations and considerations that were taken into account regarding the source material and approaches to it.

The objects of analysis in this study are referred to as prepositional constructions. Sometimes they consist of a single preposition (e.g. English *to*) and sometimes they consist of an adverbial particle and a preposition (e.g. English *up to*). The study is based on quantitative data from the *Corpus of Contemporary American English* (henceforth COCA) and *Korp* (a large corpus of Swedish texts). The prepositional constructions are analyzed from the point of view of *construal* and *image schemas*, two important constructs in the framework of cognitive linguistics.

Construal and image schemas

We can talk about, or construe, the same situation in many ways. *Construal* is a concept introduced by Langacker (2008: 55) that encompasses the degree of specificity, the focus and prominence of certain elements, as well as the perspective of the speaker in a linguistic expression. For example, both *to space* and *up to space*, describe motion to space, which is the point of focus. However, *up to* is more specific than *to*, and it adds information about where space is located relative to the speaker (above). Importantly, construing a scene in a particular way is (most often) not a conscious decision on behalf of the speaker (Langacker 2008: 88). Therefore, this article will not place any value judgments on the aptness of the prepositional constructs at hand. Rather, the analysis focuses on the way the scene is construed, with particular attention paid to which focus and perspective are assumed.

In addition to construal, the notion of *image schema* will be employed. It was developed by Lakoff (1987) and Johnson (1987) as part of their experiential approach to language (and thought). Hampe (2005: 1) summarizes their definition of image schemas as “*directly meaningful* (‘experiential’/‘embodied’), *preconceptual* structures which arise from, or are grounded in, human recurrent bodily movements through space, perceptual interactions, and ways of manipulating objects”. In other words, image schemas are neither linguistic nor visual structures. Rather, they emerge from our repeated bodily experiences and interactions with the world

around us. Through repetition they become conventionalized, while still being flexible enough to fit a wide range of contexts (Johnson 1987: 30). For this study, the CONTAINER schema and SOURCE-PATH-GOAL schema are of central importance. The CONTAINER schema “defines the most basic distinction between IN and OUT” (Lakoff 1987: 271). A prototypical CONTAINER consists of an interior and an exterior, with some sort of boundary in between. According to Johnson, this image schema emerges from our recurring experiences of containment in our everyday lives, as we “move in and out of rooms, clothes, vehicles, and numerous kinds of bounded spaces” (1987: 21). The directional particles *in* and *out*, which appear in this study, tend to have a three-dimensional container as a “prototypical landmark” (Strzelecka 2003: 148 [my translation]). As for the SOURCE-PATH-GOAL image schema, it emerges from our experiences with motion. Motion starts from a SOURCE and moves along a PATH before it reaches a GOAL. Johnson describes how “our lives are filled with paths that connect our spatial world”, both real, physical paths, but also projected or imagined PATHS (1987: 113). Spatial language, especially when movement is involved, evokes the SOURCE-PATH-GOAL image schema, and depending on how the linguistic expression is construed the focus can shift between the elements. For example, walking *from* something focuses on the SOURCE, walking *past* something focuses on the PATH, and walking *to* something focuses on the GOAL.

Corpus wrestling

Corpus-based approaches allow for fast processing of large amounts of authentic language data (McEnery, Xiao & Tono 2006: 7) and are suitable for language comparison since they allow the researcher to study how the same concept is construed in different languages (McEnery, Xiao & Tono 2006: 91). However, corpus research is contingent on the electronic resources available, which can prove challenging when working with languages which are studied to a lesser extent, such as Swedish.

Depending on the research objectives, a corpus needs to meet certain requirements. Comparative studies benefit from using a comparable corpus, which contains similar types of texts collected using the same sampling method in two or more languages (McEnery, Xiao & Tono 2006: 49). As the data is aligned from the start, the researcher can directly compare specific features and frequencies. However, the main disadvantage of comparable corpora is that they are often small and/or highly specific. This study focuses on a general feature of language (particles and prepositions) but in connection to a very specific topic (outer space). Currently, the largest comparable corpus for English and Swedish is the English-Swedish Parallel Corpus, which consists of about 1.5 million words (Altenberg & Aijmer 2008). A query for *rymden* (literally “space”) in this corpus only yields 8 results which means that

this corpus is not large enough to address the aim of this study. Instead, this study is based on investigations of two large monolingual corpora. English data was collected from COCA, which is a ~1 billion word corpus of modern (1990-2019) American English, which makes it the largest balanced corpus of its kind. In a balanced corpus, the words are evenly distributed across a range of different genres, in COCA's case spoken language, fiction, magazines, newspapers, academic publications, general web content, blog web content and TV/movies (Davies 2008). There is no large balanced Swedish corpus at present, and hence Swedish data was collected from Korp (Borin, Forsberg & Roxendal 2012) which is a large collection of sub-corpora, with more being added continuously. The interface of Korp allows users to choose which parts to query, building their own corpus within the software. Korp can be classified as a *monitor corpus*, which is a corpus that grows in size over time and can be said to “balance any need to be precise about the composition of a corpus against sheer size” (McEnery, Xiao & Tono 2006: 7). In order to align Korp somewhat with COCA, queries in Korp were restricted to sub-corpora with data from 1990-onwards. Sub-corpora of Swedish L2 data (for example, texts in Swedish written by foreign students) and regional varieties were excluded. This resulted in a corpus of about 14.2 billion words divided across 166 different sub-corpora. A full breakdown can be found in Appendix 1. It should be noted that this corpus is not balanced but skewed towards news data and colloquial internet data. In comparison to COCA, it lacks spoken data, fiction, and TV/Movies. However, despite this shortcoming, it is still the best available resource at present.

The aim of this study is to investigate the specific string “prepositional construct + *spacelrymden*”. Since the prepositional constructs can be expressed via both one word and two words, no single query will encompass all options. Instead, query terms were decided by sampling the corpora. 1000 random samples were retrieved from each corpus using the query “PREP + space/rymden”¹. The concordance lines were then manually analyzed and both one-word and two-word prepositional constructions were extracted. Table 1 shows the query terms generated by the sampling method described. However, some queries were added based on intuition, and these are marked with a *. Some of the *-marked queries did appear in the final data set but in very low numbers.

The study is concerned with motion, and the final corpus query needed to limit the results to that specific context. In order to achieve this, searches were made

¹In COCA it is possible to extract 1000 random samples from the entire corpus, whereas Korp only allows random samples from one sub-corpus at a time. Consequently, 500 random samples each were retrieved from two of the larger sub-corpora in Korp, that is, from SVT Nyheter (public service news) and Twittermix (extracts from Swedish Twitter users). These two corpora represent two different, but colloquial, registers.

Table 1. Query terms used to search COCA and Korp.

English	Swedish
<i>into space</i>	<i>in i rymden</i>
<i>up into space</i>	<i>upp i rymden</i>
<i>out into space</i>	<i>ut i rymden</i>
<i>off into space</i>	
<i>to space</i>	<i>till rymden</i>
<i>up to space*</i>	<i>upp till rymden*</i>
<i>out to space*</i>	<i>ut till rymden*</i>
<i>off to space*</i>	
<i>toward space</i>	<i>mot rymden</i>
<i>up toward space*</i>	<i>upp mot rymden</i>
<i>out toward space*</i>	<i>ut mot rymden</i>
<i>towards space</i>	
<i>up towards space*</i>	
<i>out towards space*</i>	

for the prepositional construct only when it co-occurs (collocates) with a verb. The corpora have specific collocate tools which allow for this, and the final searches were made so that only strings of the kind “VERB + prepositional construction + *spacelrymden*” were extracted. However, this does not entirely solve the challenge of motion, as the prepositional constructions can collocate with non-motion verbs (examples include *stare*, *face*, *titta* (“look”) and *längta* (“long for”). Consequently, the results were manually processed, and all instances where the prepositional constructions collocated with non-motion verbs were removed. As an example, the query for *out into space* resulted in 179 instances divided across 95 collocates. During the manual processing, 14 collocates were removed, which represented 43 of the instances.

Quantitative results

In total, after manual processing, 2035 results were retrieved from COCA, and 5229 results were retrieved from Korp. This discrepancy is expected considering the different sizes of the corpora. In order to compare the results, the frequencies have been normalized as percentages. Table 2 shows the results from COCA. The constructions have been grouped by preposition and sorted by frequency.

The two most common prepositions in the English data are *into* and *to*, which make up 99.1% of the data. They are both combined with particles but are more common as stand-alone words. *Toward(s)* is marginal in comparison at 0.9%.

Table 3 shows the results from Korp. It also contains English translations of the prepositional constructions.

The most common preposition in the Swedish data is *i*, which only appears together with particles (of which *ut* is most common). The second most common preposition is *till* and the least common is *mot*. The particle *upp* is quite common and used together with all the prepositions.

Table 2. English data from COCA. RF = raw frequency.

Query term	RF	%
<i>into space</i>	1132	55.6
<i>out into space</i>	136	6.7
<i>off into space</i>	117	5.8
<i>up into space</i>	31	1.5
Into (all instances)	1416	69.5
<i>to space</i>	573	28.2
<i>out to space</i>	12	0.6
<i>up to space</i>	8	0.4
<i>off to space</i>	7	0.3
To (all instances)	600	29.5
<i>toward space</i>	7	0.3
<i>towards space</i>	10	0.5
<i>out towards space</i>	2	0.1
toward(s) (all instances)	19	0.9
Total	2035	100

Table 3. Swedish data from Korp. RF = raw frequency.

Query term	Translation	RF	%
<i>ut i rymden</i>	out in(to)	3162	60.5
<i>upp i rymden</i>	up in(to)	1011	19.3
<i>in i rymden</i>	in in(to)	26	0.5
i (all instances)		4199	80.3
<i>till rymden</i>	to	798	15.3
<i>upp till rymden</i>	up to	91	1.7
<i>ut till rymden</i>	out to	49	0.9
till (all instances)		938	17.9
<i>mot rymden</i>	toward	61	1.2
<i>ut mot rymden</i>	out toward	23	0.4
<i>upp mot rymden</i>	up toward	8	0.2
mot (all instances)		92	1.8
Total		5229	100

In general, both languages have approximately the same number of available constructions (English has 10, Swedish has 9) of which only a few make up the bulk of the data. The following discussion deals with an analysis of the constructions through the lenses of image schema and construal, starting with the most common expression in each language.

Into and *ut i*

Into focuses on the GOAL in the SOURCE-PATH-GOAL image schema. It involves movement towards a GOAL “with a boundary and an interior” (Lindstromberg 2010: 31), where the moving object crosses a boundary to enter a GOAL. In other words, space is construed as a CONTAINER which one can enter and exit. Space does not fit into the prototypical notion of a CONTAINER, as it is infinite, without any boundaries. However, *into* does not necessarily focus on the outer boundaries of outer space but rather highlights the boundary between earth and outer space. Swedish *i* also evokes the CONTAINER schema but the direction is denoted by

the particle *ut*. *Ut* prototypically describes movement out from a CONTAINER rather than into one (Strzelecka 2003: 153). Thus, the full construction *ut i* construes not only space as a CONTAINER, but earth as well. In cases where a speaker describes motion between two containers, their perspective decides which container is more central and which one is more peripheral (Strzelecka 2003: 153). When using *ut i*, the speaker’s location (on earth) takes precedence, and earth is construed as more central. On the contrary, *into* construes space as more central, because it is the focus of the speaker’s attention. Furthermore, Strzelecka (2003: 153) also points out that *in* generally denotes motion toward a more closed-off space, whereas *ut* generally denotes motion toward a more open space. In other words, *ut i* construes space as both less central (to the speaker) and more open.

Both *into* and *ut i* construe motion between earth and space as an IN/OUT motion, which is interesting considering that for vertical motion, *up* and *down* usually take precedence. Strzelecka even goes as far as saying that “[w]ithin the physical domain there are practically no particle verbs with *in* and *ut* which refer to vertical motion” (2003: 205 [my translation]). Either motion from earth to space is a case where the notion of containment takes precedence over vertical movement, or it is not necessarily perceived as vertical movement at all. From a perspective of standing on earth, you always need to move upwards to reach space. However, if the perspective is zoomed out we know that motion away from earth is not unidirectional (an astronaut from the northern hemisphere and an astronaut from the southern hemisphere can move in opposing directions but still both reach outer space).

It should be noted that *into* also combines with particles. In these cases, *into* denotes motion into a space-CONTAINER, and the particle describes where that container is located. The constructions provide more information on the perspective of the speaker. *Out* functions like *ut*, shifting the focus from GOAL to encompass both SOURCE and GOAL. *Up* instead highlights the vertical direction. According to Lindstromberg (2010: 55) the prototypical meaning of *off* is loss of support, or “separation from (the surface of) a supporting landmark”. In the construction *off into space*, *off* means separation from the surface of the earth, and thus focuses on the SOURCE and first part of the PATH. No Swedish particle has this specific meaning.

To and *till*

In constructions with *to* and *till*, the GOAL is particularly salient as the preposition specifies moving towards and reaching “the endpoint of a path” (Lindstromberg 2010: 30). Looking to the notion of CONTAINMENT, *to* and *till* are often pragmatically understood as reaching and entering a GOAL, rather than just going up to the boundary of it (Teleman, Hellberg & Andersson 1999: 699).

As such, going *to space* or *till rymden* can also be understood as going into space. Lindstromberg (2010: 31) argues that, in English, the choice between *to* and *into* has to do with “the scale of our mental image” of the GOAL. *To* is most often used when the distance to the destination is longer. The closer we get to the GOAL, the more apparent the physical boundary (and need to cross it) becomes, and people are more likely to use *into*. For example, we go *to* a lake, but we dive *into* the water. The data in this study shows that both *into* and *to* are used to describe motion from earth to space, even though we all should share the same experience of its scale. However, Lindstromberg (2010) clearly states that it is the scale of the *mental image*, rather than our physical experience, that tends to be the determining factor. The boundary between earth’s atmosphere and space is not a clear one, rather, it is a gradual decline in gravity. The lines that we impose in zoomed-out pictures and illustrations are human constructs, rather than physical realities, used to describe (for example) different layers of the atmosphere. Thus, these lines may or may not be a part of people’s mental image of space, depending on which perspective they assume. This ambiguity might explain why both options (*into* and *to*) appear in the data.

Up and upp

Up and *upp* relate to movement along a vertical axis, and according to both Lindstromberg (2010: 189) and Strzelecka (2003: 213) this vertical axis is connected to gravity. Strzelecka (2003: 213) claims that *upp* always and unambiguously refers to movement along a vertical axis, with the ground as the canonical starting point of the motion. *Up* and *upp* prototypically describe an upwards movement, and they are primarily GOAL-focused even though the SOURCE is still present, albeit implicitly. However, the meaning of these particles is somewhat complicated in the context of space. When an upward motion reaches high enough, *up* and *upp* stop having any discernible meaning since gravity is no longer a factor. In other words, the perspective is given when these particles are used, and we can assume that the speaker is standing on the surface of the earth. The focus is thus primarily on the initial part of the PATH towards space.

As for CONTAINMENT, these particles do not invoke that notion on their own, only when combined with *into* or *i*. Strzelecka (2003: 216) explains that the sky can be construed as a CONTAINER, which is open at the bottom, similar to how we construe water masses as CONTAINERS with an opening at the top (the surface). Going *up into* or *upp i rymden* thus construes space as a (less prototypical) container that one can enter from below.

Finally, it should be noted that *upp* is more common in Swedish (appearing in 21.2% of the constructions) than *up* is in English (appearing in 1.9% of the constructions). This construction, and the perspective that it entails, is thus more common in Swedish than in

English. However, what this depends on is outside the scope of this study. It is possible that *off* is filling the same niche in English since that also primarily focuses on the initial parts of the PATH. However, one would need to consult the context in more detail to investigate that possibility.

Toward(s) and mot

The least common constructions in both languages are *toward(s)* and *mot*, and they are the only constructions in the data set that primarily focus on the PATH. They describe motion in a particular direction (towards a GOAL), but the moving object does not necessarily reach said GOAL. In that sense *towards* and *mot* both denote an “unrestricted action” (Teleman, Hellberg & Andersson 1999: 791), an action that is not completed. Furthermore, like *off* and *upp*, *toward(s)* focuses on the initial part of the PATH, before reaching (and crossing into) space. Considering that most intentional movements toward space reach their destination, this focus on the PATH essentially backgrounds the GOAL.

Conclusions

This study has shown that there are several different ways of using prepositional constructions to construe motion from earth to space. Most constructions focus on the GOAL, but the SOURCE is often implicit. In constructions where the PATH is the primary focus, the first part of the PATH is typically highlighted. As for CONTAINMENT, both space and earth can be construed as containers that one can move in between.

Generally, the prepositional constructions show that the speaker can assume different perspectives, or ways of construing, the same situation. The most common constructions in both languages, combinations of particles with *intoli* and *totill* assume a more zoomed-out perspective, as they invoke the full range of motion from earth to space. This is the way that we are familiar with space (and earth) being pictured and illustrated. On the contrary, in constructions with the particle *uplupp* or the preposition *toward(s)/mot* the perspective is clearly that of the speaker’s physical location. As for Swedish and English, they differ when it comes to the most conventionalized expression in each language: English assumes a more space-focused perspective whereas Swedish assumes a more earth-focused perspective. The other prepositions are largely similar in both languages (with *to* and *till* being the second most common and *toward(s)* and *mot* being the least common). The use of particles also differs between the languages as *upp* and *ut* (and the perspective they entail) are more common in Swedish than in English. The reasons for these differences in language and perspectives are an interesting line of inquiry for future research.

As for methodology, this study has built a foundation for further research by combining corpus linguistic analyses with theory. However, corpus data alone does not

reveal why certain constructions are favored over others and cannot account for differences both within and between the languages. In order to reach a deeper understanding of this, one could employ either discourse analytical approaches or psycholinguistic surveys in the future. However, this study has provided initial insights into tendencies and prototypical usage within each of the languages.

Author Bio

Kajsa Törmä is a doctoral researcher in English linguistics at the department of language studies at Umeå University. Her thesis project concerns how we conceptualize outer space as a physical place and how we utilize our empirical knowledge, embodied cognition, and imagination to do so.

Electronic corpora

- Davies, Mark (2008). *The corpus of contemporary American English (COCA)*. <https://www.english-corpora.org/coca/> [October 7 2022].
- Lars Borin, Markus Forsberg, Johan Roxendal (2012). Korp – the corpus infrastructure of Språkbanken. *Proceedings of LREC 2012. Istanbul: ELRA*, 474–478. <https://spraakbanken.gu.se/korp/> [October 7 2022].
- Altenberg, Bengt and Karin Aijmer (2000). The English-Swedish parallel corpus: A resource for contrastive research and translation studies. In: Mair, Christian & Hundt, Marianne (eds.) *Corpus linguistics and linguistic*

Theory. Atlanta: Rodopi. <https://spraakbanken.gu.se/resurser/espc> [October 7 2022].

References

- Hampe, Beate (2005). Image schemas in cognitive linguistics: Introduction. In: Hampe, Beate & Grady, Joseph E. (eds) *From perception to meaning: Image schemas in cognitive linguistics*. Berlin: New York: Mouton de Gruyter.
- Johnson, Mark (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago: University of Chicago Press.
- Langacker, Ronald (2008). *Cognitive grammar: A basic introduction*. Oxford: Oxford University Press.
- Lakoff, George (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
- Lindstromberg, Seth (2010). *English prepositions explained*. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- McEnery, Tony; Tono, Yukio & Xiao, Richard (2006). *Corpus-based language studies: An advanced resource book*. London: Routledge.
- Strzelecka, Elzbieta (2003). *Svenska partikelverb med "in", "ut", "upp" och "ner": En semantisk studie ur kognitivt perspektiv*. (diss.) Uppsala University.
- Teleman, Ulf & Hellberg, Staffan & Andersson, Erik (1999). *Svenska akademiens grammatik: 2 Ord*. Stockholm: Svenska akademien
- Tyler, Andrea & Evans, Vyvyan (2003). *The semantics of English prepositions*. New York: Cambridge University Press.

How to cite this article: Kajsa Törmä. 2022. Going to space: A cross-linguistic study using electronic corpora. *Kulturella Perspektiv*, vol. 31. Theme: Sources, s. 1–6.

Submitted: 25 April 2022 **Accepted:** 24 October 2022 **Published:** 23 December 2022

Copyright: © 2022 The author(s). This is an Open Access article, distributed in accordance with Creative Commons, license CC-BY 4.0, which permits unlimited use, distribution, and reproduction in all media formats, granted that original author(s) and source are duly attributed. See <http://creativecommons.org/licenses/by/4.0/>.

Kulturella Perspektiv – Svensk etnologisk tidskrift, ISSN 1102-7908, is a peer reviewed Open Access journal, published by the association Föreningen Kulturella Perspektiv.

OPEN ACCESS 