

Through and Through: Corpus-Based Semantic Analysis of a Spatial Preposition

Brandon Bennett   

School of Computing, University of Leeds, UK

Maria M. Hedblom   

Department of Computer and Systems Sciences, Stockholm University, Sweden

Department of Computing, Jönköping University, Sweden

Abstract

We present a corpus-based analysis of sentences containing the highly polysemous word “through” and establish a framework for categorising sentence constituents in relation to this preposition. Using the annotated corpus, together with ideas from the fields of Lexical Semantics and Knowledge Representation, we specify a typology of *figure*, *path-verb*, and *ground* terms, characterised by the *affordances* they provide for participation in “through” relationships. We find that the mode of affordance provided by the *ground* is often ambiguous: many types of object can provide multiple modes of grounding for a “through” relationship, associated with different geometrical conditions between figure and ground. However, this ambiguity can be explained and often resolved or reduced by considering compatible combinations of figure and ground affordances in relation to the path.

2012 ACM Subject Classification Computing methodologies → Knowledge representation and reasoning; Computing methodologies → Natural language processing; Computing methodologies → Spatial and physical reasoning

Keywords and phrases Preposition, Polysemy, Semantics, Spatial Relation, Ontology

Digital Object Identifier 10.4230/LIPIcs.COSIT.2026.15

Category Short Paper

1 Introduction

Prepositions are an essential component of English and many other natural languages. They form a relatively small (“closed”) class of words that combine with other words to describe fundamental spatial and temporal relationships between objects in the world. A preposition that displays a very wide range of senses is “through.” While usually regarded as primarily a spatial preposition, it has many uses in describing temporal, causal, informational and other abstract relationships. It can occur in several grammatical roles and interacts in complex ways with associated verbs as well as both physical and abstract objects. Even when confining our attention to spatial descriptions, we can contrast cases such as: “*The boy ran through the tunnel*”, “*A rope ran through the tunnel*” and “*Water ran through the tunnel*”; as well as “*She ran a comb through her hair*”, “*She threw the comb to him through the window*” and “*Moonlight shone through the window.*”

One may feel that there is some essential “throughness” that unites the situations. According to the *Image Schema* approach, the senses of prepositions, correspond to fundamental but highly adaptable patterns in our experience. [7, 5]. Each use of “through” conveys a particular range of possible spatio-temporal relationships holding between the subject (the *figure*) and the prepositional object (the *ground*) but the geometrical relationships involved can be very different. We aim to illuminate this diversity by analysing a corpus of sentences containing “through” and developing a framework for describing the roles of sentence constituents in such constructions. Drawing on ideas from lexical semantics and knowledge representation, we introduce a typology of *figure*, *ground*, and *path-verb* expressions, characterised by the affordances they contribute to a through relationship.



© Brandon Bennett and Maria M. Hedblom;
licensed under Creative Commons License CC-BY 4.0

17th International Conference on Spatial Information Theory (COSIT 2026).

Editors: Sabine Timpf, Gabriele Filomena, Armand Kapaj, Rui Zhu, Nicholas Giudice, and Ed Manley;
Article No. 15; pp. 15:1–15:7



Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

2 Foundations and Previous Work

The massive variety of meanings conveyed by the word “through” offers an interesting study of the phenomenon of *polysemy*, where a word can be used in many different but related senses [4]. Because “through” can be used in so many different ways, it is difficult to determine whether two different examples can be considered as two cases of a general form of situation, or whether they exemplify different senses (See Section 4). Hence, we felt the need for an organising framework. The influential work of Talmy [9] construes the semantics of motion verbs and event descriptions as decomposable into the following interacting components:

Figure: the entity whose location or configuration changes,

Ground: the reference entity with respect to which the change is characterised,

Motion: the occurrence of displacement or maintenance of location over time,

Path: the geometric or topological structure of that displacement (direction, boundary-crossing, traversal),

Manner: the way in which the Motion is carried out (running, sliding, limping, crawling),

Cause: whether the Motion is externally induced (push, throw, carry).

In English Motion and often Manner are encoded in the verb, while properties of the Path are expressed by a preposition or particle (e.g. *run into*, *fall off*). For present purposes, this division is central, since the preposition “through” contributes a specification of Path that interacts systematically with the Figure, the Ground, and the verb.

A complementary perspective is provided by Pustejovsky’s *Generative Lexicon* theory [8]. Rather than treating polysemy as a fixed list of discrete senses, the theory proposes that lexical items encode rich internal structure from which context-sensitive interpretations are generated. Composition is guided by a system of types and by mechanisms of coercion that allow operator expressions (verbs, adjectives, prepositions) to adjust the semantic type of their arguments when required. Hence, the meaning of a construction of the form “ α ϕ ’s ‘through’ β ” is not obtained by simply selecting a stored sense of “through”, but emerges from interaction between (i) the semantic type and internal structure of the Figure α , (ii) the event structure and path-contributing properties of the verb ϕ , (iii) the structural and topological properties of the Ground β , and (iv) a family of constraints associated with “through”. The resulting interpretation is the outcome of a compatibility resolution among these elements.

A central component of the Generative Lexicon is the *qualia structure*, which characterises an object along four dimensions: *formal* (what kind of thing it is), *constitutive* (its material or parts), *telic* (its function or purpose), and *agentive* (how it comes into being). For spatial expressions, such structure is particularly relevant because it encodes affordances: a tunnel affords passage, a wall resists it, a net permits partial penetration while constraining motion. Many uses of “through” appear to depend not merely on geometry but on such functionally grounded properties of the Ground and Figure. The qualia framework therefore, provides a principled way of modelling why certain objects are naturally construed as passable, penetrable, or traversable.

3 Our Methodology

Our methodology combines intuitive judgements about possible meanings of the word “through” with formal analysis from Lexical Semantics and logical theories of spatial and temporal relations. For our study we used the fiction texts of the Brown Corpus [6],¹ a widely

¹ The sentences of the corpus, the Python code used in the analysis, the semantic annotations, and some simple functions for displaying annotated sentences, category structures and frequency counts are available via *GitHub* (made available on acceptance to maintain anonymity).

used dataset suitable for reproducible analysis. The fiction section contains 275 occurrences of the word “through”. Although not comprehensive, the relatively small but spatially rich sample allows manual annotation of every instance.

For each occurrence we recorded (i) the associated verb, (ii) the figure, the entity whose participation in the relation is conceptually central, and (iii) the ground, the entity or region introduced by “through”. This stage required careful hand annotation because syntactic position was often misleading: the figure is not always the grammatical subject and may sometimes be implicit. In parallel we classified grounds into salient semantic types (see Section 4.3) and began a corresponding classification of verbs into event types (see Table 2).

Lightweight Python scripts were used for annotation and analysis but not for the semantic analysis itself. A language model (**ChatGPT-5.2**) was also used to assist with coding and occasionally to suggest likely sentence components or provisional classifications. Its output was treated only as indicative and was always checked against the original sentences.

Rather than beginning with a fixed ontology, we moved iteratively between the corpus and the developing classification. New examples often prompted additions or refinements to the category system. At the same time we sought generality and definability, identifying characteristics that applied across many cases and appeared amenable to precise definition. In this way corpus evidence, linguistic intuition, and type-based reasoning were used together, producing an analysis that is both empirically grounded and systematically organised.

4 Analysis

One may initially consider “through” to be a binary spatial relation. However, consideration of examples of its use reveals that its meaning often depends on an associated verb. Hence, the expression of “through” relationship can be decomposed into the following elements:

*Figure, Path-Verb, **through**, Ground.*

4.1 Figures and their Types

Many types of entities can be the Figure in a “through” relationship description (see Table 1). Different types of physical objects provide different affordances, which influence the ways they can participate in “through” relationships. Hence, we have subcategorised PHYSICAL-OBJECT with respect to salient affordances as follows: ELONGATED-OBJECT (“sticks”, “rifle muzzles”, “bolo”, “stake”), ANATOMICAL-OBJECT (“skull”, “hair”, “body”), PROJECTILE (“bullet”, and two pronouns with “bullet” implied), AGGREGATE (“snowflakes”), PHYSICAL-PATH (“alleyway”), BUILDING (“cottages”), LIQUID-OBJECT (“river”), FILAMENTED-INSTRUMENT (“comb”), PHYSICAL-OBJECT (“basket”), PHYSICAL-BARRIER (“revetments”), *Mixed-type Figure* (“guns and fists”).

The Figure contributes its trajectory (whose properties may also be described by the verb) and properties such as self-motivation, elongation, fluidity, hardness, or sharpness.

4.2 Verbs Associated with “Through”

We examined the verbs occurring with through in every sentence of the corpus and grouped them into categories (Table 2). Our categorisation is intended to cover all categories in which the verb can function. For instance, “run through” appears in the corpus only in the sense we call ITERATIVE-SCAN (as in “run through the info”). However, run has several senses, and is therefore also included in the SELF-MOTION, STATIC-ROUTE, and CAUSED-MOTION categories.

■ **Table 1** Figure types in the “through” corpus.

Type	Count	Sub-Types
PERSON	174	Includes plurals, 11 cases implied/unreferenced.
PHYSICAL-OBJECT	20	‘sticks’, ‘hair’, ‘cottages’
ANIMAL	5	3 ‘horse’, ‘nymph’ (insect), ‘birds’
VEHICLE	4	‘jeep’, ‘van’, ‘truck’, ‘wagon’
LIGHT	11	5 ‘light’, 2 daylight, ‘moonlight’, ‘colours’, ...
INFORMATION	10	‘lyrics’, ‘news’, ‘messages’, ‘smile’, ‘images’, ...
EMOTION/SENSATION	8	‘twinges’, ‘loneliness’, ‘pity’, ‘chill’, ‘anger’, ...
MENTAL-ENTITY/CAPACITY	4	3 ‘mind’, ‘thought’
GASEOUS-SUBSTANCE	8	2 ‘air’, 2 ‘smoke’, ‘scent’, ‘breath’, ‘gases’, ‘wind’
LIQUID-SUBSTANCE	5	5 ‘water’, ‘sweat’
EVENT/PROCESS	5	‘life’, ‘flight’, ‘childhood’, ‘deal’, ..
SOUND	5	4 ‘sound’, ‘voice’
Other types with only 1 example	7	SOLID-SUBSTANCE, PHYSICAL-MOVEMENT, COLOUR, PLACE, GEOM-FIGURE, AUDIO-QUALITY, IMPLIED-GOAL

■ **Table 2** Verb types with respect to *through* relationships. Verbs may appear in multiple categories due to polysemy.

Verb Type	Count	Items
SELF-MOTION	114	bolt, climb, come, crawl, cross, disappear, dive, drive, ...
STATIC-ROUTE	43	go, lead, lie, pass, run, thread
TRANSMISSION	36	carry, come, enter, flash, glow, pass, reach, show, ...
COMPLETION	36	come, get, go
PROPAGATION	28	course, disappear, filter, flood, glow, ripple, run, seep, ...
PERCEPTION	28	gaze, glance, hear, look, peek, peer, read, see, watch
CAUSED-MOTION	22	bring, carry, coax, drag, draw, force, hand, hurl, puff, ...
PENETRATION	17	crash, cut, drive, jab, pierce, rake, scuff, shoot, tear, wear
ITERATIVE-SCAN	15	leaf, pick, plod, run, search, seek, sift
ENDURANCE	8	carry, live, nurse, stay, suffer, sustain
CAUSAL-MEDIATION	5	acquaint, do, know, smooth, start

4.3 Ground Descriptors and their Types

We found great variety in affordances supplied by the Ground in relation to “through” and found them much harder to describe and categorise than those associated with the Figure. It was often difficult to identify genuinely different modes of affordance among the many possible geometries that an instance “through” relation may take. By looking for conceptual generalities among the particular cases we arrived at the following typology:

- CONTAINER – an enclosed region that can be entered and exited.
 - MULTI-APERTURE – multiple distinct openings enabling traversal across it.
 - SINGLE-APERTURE – essentially one entry/exit opening.
 - STRUCTURED – internally partitioned into compartments.
- FRAME – a bounded opening defined by a surrounding object rather than a volume.
- APERTURE – an opening in a barrier separating two sides.
- APERTURE-PROXY – an object type used to refer to an associated aperture (e.g. a door or curtain).

- BARRIER – a surface or sheet obstructing movement between regions.
 - GAPPY-BARRIER – barrier containing holes or spaces permitting passage.
- AGGREGATE – a collection of discrete objects forming an extended obstruction or medium.
- ENVIRONMENT – an extended surrounding spatial region.
 - ENV-2.5D – surface-dominated terrain with horizontal extent.
 - ENV-3D – volumetric enveloping medium (air, fog, water, darkness).
 - CONDUIT – elongated environmental channel guiding motion (street, canyon).
- INFORMATION-CONDUIT – supports communication or perceptual access.
- BODILY-MEDIUM – parts of an organism functioning as a penetrable or perceptual medium.
- TIME-PERIOD – an extended temporal interval conceptualised as traversable.
- ENACTMENT – a sequence of actions forming a coherent episode.
- CAUSAL-MEDIUM – an abstract condition or circumstance producing an outcome.
- INFORMATION-SOURCE – a structured body of information examined sequentially.
- AUDIO-LAYER – a surrounding field of sound.
- FIGURE-2D – a purely geometric region.
- Additional Qualifier Types: UPRIGHT, GROUNDED (anchored to the ground), FILAMENTED (grass, fingers), LIGHT-TRANSMITTING, LOS-TRANSMITTING (apertures affording a line of sight).

4.3.1 Ground Vocabulary Type Lexicon

We use the terms above to define an affordance lexicon, which associates nouns occurring as the Ground in a “through” phrase with a boolean combination of categories and qualifiers. Of particular interest are those whose types are *disjunctive* (with $\vee$ in the type specification). These are cases where the ground object presents more than one mode of affordance with respect to potential “through” relationships. We found that anatomical objects provide especially interesting diversity in their affordances.

Informal grouping identified families of ground types: Containers, Apertures, Barriers and aggregates, Vegetated environments, Environment situated conduits, 3D environments, Anatomical objects, Time and Informational media. Each of these families provides a distinct range of affordance types in association with “through”; and, moreover, even these refined types could often provide several alternative affordances depending on their context. Complex affordance types are represented as follows:

Barriers and aggregates

<i>grille, veil</i>	GAPPY-BARRIER
<i>curtain</i>	BARRIER $\vee$ APERTURE-PROXY
<i>fence</i>	(GROUNDED $\wedge$ UPRIGHT $\wedge$ BARRIER) $\vee$ (GROUNDED $\wedge$ UPRIGHT $\wedge$ GAPPY-BARRIER)
<i>trees</i>	FILAMENTED $\vee$ (GROUNDED $\wedge$ UPRIGHT $\wedge$ AGGREGATE)
<i>crowd, traffic, grass</i>	GROUNDED $\wedge$ UPRIGHT $\wedge$ AGGREGATE

Anatomical Objects (often construed as Bodily Media)

<i>body</i>	BODILY-MEDIUM
<i>hand</i>	BODILY-MEDIUM $\wedge$ FILAMENTED
<i>finger</i>	BODILY-MEDIUM $\wedge$ FILAMENTED $\wedge$ INFORMATION-CONDUIT
<i>nose</i>	BODILY-MEDIUM $\wedge$ MULTI-APERTURE $\wedge$ CONTAINER
<i>eye</i>	BODILY-MEDIUM $\wedge$ LOS-TRANSMITTING $\wedge$ INFORMATION-CONDUIT

Typically, the *head* noun of a phrase yields the affordances of its referent. But this is not always the case. For example, “opaque glass” lacks the light-transmission property normally afforded by “glass”. Hence, we use an additional phrase-level lexicon, so that affordances of objects described by multi-word phrases are more accurately captured.

5 Sense Annotations and Affordance Combinations

We now present some of the outputs from the main display script that we use to view and evaluate our annotations. To give a general flavour, we start by just giving the first two, and give one more example that we consider to be of interest.

0. "Rachel wore a smart hat and, because she had been warned recently about smoking, puffed at her cigarettes **through** a long ivory holder stained with lipstick."
 FIGURE: 'she' (PERSON)
 PATH VERB: 'puffed' (puff) (CAUSED-MOTION)
 GROUND: 'holder' ((SINGLE-APERTURE | MULTI-APERTURE) CONTAINER)

1. "At 2130 hours they had passed **through** the barbed wire at the point of departure."
 FIGURE: 'they' (PERSON)
 PATH VERB: 'passed' (pass) (TRANSMISSION | SELF-MOTION | STATIC-ROUTE)
 GROUND: 'barbed wire' (GROUNDED UPRIGHT FILAMENT BARRIER)

6. "He thought how this dainty, fragile older woman threading her way **through** the streets of Westminster on a day in June, enjoying the flowers in the shops, the greetings from old friends, but never really drawing a deep, passionate breath, was so like himself."
 FIGURE: 'this dainty, fragile older woman' (PERSON)
 PATH VERB: 'threading' (thread) (SELF-MOTION)
 GROUND: 'day' (TIME-PERIOD)

The final example is interesting because of the error in the *ground* annotation, which we subsequently corrected. Here, the error is plain to see because the affordances of the *figure* and *verb* do not fit with those of the *ground*.

5.1 Distinct Spatio-Temporal Relationships

We now turn to the question of what are the spatial conditions described by a “through” relationship. We are currently working with the following typology of “through” relations:

- APERTURE-CROSSING: Passage through a local opening in an otherwise blocking surface. (e.g. through the door, window, gap, keyhole)
- APERTURE-TO-APERTURE-TRAVERSAL: Passage through a local opening in an otherwise blocking surface. (e.g. through the door, window, gap, keyhole)
- CONDUIT-TRAVERSAL: Guided movement within an extended enclosed channel. (e.g. through the tunnel, pipe, cave passage)
- DESTRUCTIVE-BARRIER-PENETRATION: Penetration of resistant barrier (e.g. bullet through the wall, nail through the board, knife through fabric, brick through windowpane)
 Salient sub-types are: PIERCING and SMASHING/BREAKING
- CONTINUOUS-MEDIUM-TRAVERSAL: Motion within a yielding surrounding substance. (e.g. through water, air)
- AGGREGATE-OBJECT-TRAVERSAL: Passage through a densely occupied region. (e.g. through the crowd, traffic, forest)
- AGGREGATE-BARRIER-CROSSING: Crossing a barrier constituted by an aggregate.
- INTERIOR-REGION-TRAVERSAL: Movement across a bounded bounded volume. (e.g. through the room, building, city)

- **THREADED-TRAVERSAL:** Guided passage along a shaped interior path. (e.g. thread through the needle, rope through the pulley)

Our ongoing work is targeting two goals: (1) defining the geometrical conditions of spatio-temporal relationships corresponding to the different senses of “through”; (2) showing how our typology of the Figure, Verb and Ground types can be used to resolve or significantly reduce the ambiguity of “through” in any given context; and hence determine the spatio-temporal conditions it implies. Though we do not have space to present them here, we have already specified representations of several of the above relations in a spatio-temporal logic, making use of some ideas presented in [2] and [3], and of similar style to that presented in [1]. From our annotation outputs (see Section 5) we have worked out mappings to spatial conditions for many cases in our corpus.

6 Conclusion

Due to its huge range of senses, the meaning of “through” is resistant to straightforward classification and formalisation. Nevertheless, we believe that examining deeper semantic structure in terms of types and affordances of *figure*, *path-verb* and *ground* provides a promising route to a systematic account, that could support resolution of sense ambiguities of this highly polysemic word over a wide range of contexts. Although focused on a single word, we believe that our meticulous categorisation methodology can be applied to other words, and and may shed light on the phenomena of polysemy. Our preliminary findings are consistent with the view that meanings of polysemic words are resolved by accommodation of affordances between interacting semantic components within a sentence. Specifying mechanisms for this resolution is a remaining aim for the further development of our analysis.

References

- 1 Brandon Bennett, Vinay Chaudhri, and Nikhil Dinesh. A vocabulary of topological and containment relations for a practical biological ontology. In John Stell, Thora Tenbrink, and Zena Wood, editors, *Spatial Information Theory, 11th International Conference, COSIT 2013*, LNCS, Scarborough, UK, 2013. Springer. doi:10.1007/978-3-319-01790-7_23.
- 2 Anthony G Cohn, Brandon Bennett, John M Gooday, and Nick Gotts. Rcc: a calculus for region based qualitative spatial reasoning. *GeoInformatica*, 1(3):275–316, 1997.
- 3 Nicholas Mark Gotts. How far can we ‘C’? Defining a ‘doughnut’ using connection alone. In *Principles of Knowledge Representation and Reasoning*, pages 246–257. Elsevier, 1994.
- 4 Stefan Th. Gries. Polysemy. *Cognitive linguistics: Key topics*, pages 473–490, 2019.
- 5 Maria M. Hedblom, Dagmar Gromann, and Oliver Kutz. In, Out and Through: Formalising some dynamic aspects of the image schema Containment. In *Proc. 33rd Annual ACM Symposium on Applied Computing*, pages 918–925, Pau, France, 2018. doi:10.1145/3167132.3167233.
- 6 Henry Kučera and W. Nelson Francis. *Computational Analysis of Present-Day American English*. Brown University Press, Providence, RI, 1967.
- 7 George Lakoff and Mark Johnson. *Metaphors We Live By*. University of Chicago Press, 1980.
- 8 James Pustejovsky. *The Generative Lexicon*. MIT Press, Cambridge, MA, 1995.
- 9 Leonard Talmy. *Toward a Cognitive Semantics*. MIT Press, Cambridge, MA, 2000.