Situated interaction with a smart environment

Tenbrink, Thora

KI - Künstliche Intelligenz

DOI:
10.1007/s13218-017-0495-7

Published: 01/08/2017

Peer reviewed version

Situated interaction with a smart environment: Challenges and opportunities

Thora Tenbrink

Bangor University, Wales, UK

+44 1248 382263

t.tenbrink@bangor.ac.uk

http://knirb.net

Interacting with a smart environment involves agreeing on what to do when, based on a joint understanding of where things and people are or where they should be. Face-to-face interaction between humans, or between humans and robots, implies clearly identifiable perspectives on the environment that can be used to establish such a joint understanding. A smart environment, in contrast, is ubiquitous and thus perspective-independent. This paper reviews the implications of this situation in terms of the challenges for establishing joint spatial reference between humans and smart systems, and presents a somewhat unconventional solution as an opportunity.

Keywords: Perspective, communication, spatial reference, common ground

"Where are my pills?" asks an elderly resident of a smart home, and expects the ubiquitous assistive system not only to know where the pills are but also to be able to help the user find them. On the face of it, this is a simple everyday situation: a mere pointing gesture or a helpful action would be enough for a human assistant who is physically present, sharing the same environment and knowledge concerning the nature of the object in question. However, in the absence of such easy solutions, considerable challenges arise. First, user and system need to establish joint reference as a pre-requisite for correct access to its knowledge base. That is, the system needs to understand exactly which item the user is asking for, and identify its specification in the database. Second, the system's knowledge base needs to provide a link between the object referred to and a place in the environment. This will happen in whatever formats the system works with; for instance, the item in question might have an ID tag in the database, which is cross-linked with a spatial ID in another database that represents the spatial setting. Third, the spatial information needs to be conveyed to the user in a way the user can handle. Simply offering the spatial ID from the database will not be sufficient, and the ubiquitous assistance system will not be

capable of providing direct access to the object or pointing to it. Visual representations pose additional challenges, and they may not be feasible in the context of a complex smart home environment that includes many small movable objects. In such a case, the relevant knowledge may be need to be conveyed through natural language, potentially supplementing other modes (Srimal & Jayasekara, 2017).

There are a number of ways in which language can be used to transfer spatial knowledge – and each of them comes with its own challenges. If the system has access to the user's previous actions, it could remind them by saying "You took them this morning during breakfast", thus entirely avoiding spatial reference, and leaving it to the user to make the inferences required to find the pills' current location. However, this presupposes a substantial, detailed, and complete knowledge base that ensures that no intervening actions could have led to the pills being located somewhere else. Also, the user would have to remember their morning actions sufficiently to infer the current location of the pills. Alternatively, the system could offer route directions to the user, guiding them towards a place where the user can easily spot the pills. This requires detailed knowledge about where exactly the resident is in relation to the pills, and gradual updating as they move through the environment. Presumably the simplest method, then, is to describe the object's location, as this requires no other knowledge than where the pills are currently located. This strategy for conveying spatial information is widely used in everyday life, and thus poses a perfectly natural way of communicating between humans. However, at a closer look it comes with a range of challenges for computational systems.

This paper will discuss this kind of assistance scenario by looking at how humans establish joint spatial reference with an interaction partner, and identifying the challenges and opportunities for ubiquitous systems. Notably, such systems lack a central element present in all human or humanoid interactants: an intrinsic point of view. However, they may be better suited to make use of a conceptual framework that is less widely used by humans: anchoring reference in a ubiquitously accessible, *absolute* reference system.

Spatial language

Although the linguistic repertory for establishing spatial relationships is wide and varied, it comes with a range of systematic patterns. As detailed by Talmy (2000), Bateman et al. (2010) and others, spatial language is schematic and abstract, allowing for a wide range of relationships to be expressed by a limited number of linguistic expressions. This is true, to different degrees, for all word categories (or parts of speech) that are capable of conveying spatial information. A noun like *box,* for instance, implicitly conveys a certain shape and geometric contour, while leaving size and volume underspecified. Prepositions are even more flexible; their precise interpretation in context will have to be derived from situational factors such as the nature of the objects that are related to each other by a preposition. For instance, in *the ceiling above the floor* the spatial relation between ceiling and floor is conveyed schematically by the term *above*, but there is no specification of distance. World knowledge about the typical location of ceilings relative to floors can enhance this type of spatial information; humans will have a fairly realistic idea of the range of possible distance between ceiling and floor.

Indeed, speakers frequently – and automatically – add such world knowledge in whenever spatial prepositions are used. The preposition *in*, for instance, arguably suggests a geometric relationship of containment, which is complete in the case of *the biscuit in the box,* but not in the case of *the flower in the vase* (Herskovits, 1986). In the latter case, it is typically only the stem of the flower that is actually contained in the vase. Human speakers will naturally make the relevant inferences and assume a natural state of affairs, without needing to specify details by saying *the stem of the flower is in the vase*. Such a statement would, in fact, be perceived as rather odd in most contexts.

To account for such effects in the use of the spatial prepositions *in* and *on*, Coventry & Garrod (2004) proposed the notion of dynamic-kinematic extra-geometric routines. In the above example, the *flower* can be represented as being located *in* the vase because it will move along with the vase if the vase changes its location. This kind of dynamic-kinematic routine therefore accounts for the concept of containment even if the most salient part of the object (the blossom) is not contained in the vase at all.

Schematic and functional effects like these pertain to the verbalization of all spatial relationships to some extent, and they are extremely context- and world-

knowledge dependent. Such effects pose tremendous problems to automatic systems that do not have access to the considerable richness of a human's experiential knowledge. Each noun denoting a spatial object (such as the *box* mentioned above) has experientially based connotations, and the use of prepositions typically presupposes knowledge of experientially possible or likely routines.

To some extent, object features can be ignored by treating any object as point-like rather than considering its actual contour and extent, and relying entirely on the schematic spatial information conveyed by prepositions, based on the fact that they are the most versatile linguistic expressions available. This strategy has often been adopted in spatial reasoning approaches (Freksa, 1992; Moratz & Ragni, 2008, Rodrigues, Santos, & Lopes, 2016), with the aim of developing generic strategies for generating spatial reference by automatic systems across situational contexts. The main prerequisite for this kind of strategy is to understand the semantic contribution made by each of the spatial prepositions (or preposition-like structures, such as the phrase *in the front of*). This can be treated either independently of the contextual details conveyed by experiential knowledge of objects and their features and functional interrelationships, or by identifying generic functional features that can be implemented in an adaptive system (Guadarrama et al., 2013). In this regard, different types of prepositions pose different types of challenges, which can be briefly sketched as follows (see Tenbrink, 2007 for further details of spatial term categories).

Topological terms

This class encompasses expressions of neighborhood or contiguity, such as *at, on*, and *in.* The terms are extremely frequent in everyday language for the description of spatial locations between objects of any type. The main challenge for automatic systems associated with this class is the issue of determining the scope of proximity, which can be based on absolute and relative concepts (Kelleher, Kruijff, & Costello, 2006). It is particularly true for the class of topological terms that object features and functions are decisive for their use in context (Coventry & Garrod, 2004). In addition to the effects outlined above, speakers tend to have implicit expectations as to when topological terms can be meaningfully used. For instance, although a tablecloth is geometrically located *on the table,* a more usual

way of expressing its location would be to say that it *covers the table* or that it's *over the table.* The utterance *The tablecloth is on the table* appears to suggest that it's lying there folded up and perhaps ready for use, but not actually in use. Such associations are independent of the actual spatial relationship suggested by the terms involved (the prepositions *on* and *over,* and the verb *covers*), but are experientially conveyed. Also, they are context- and object-specific to a very high degree, posing considerable challenges for implementation in any automatic system; this accounts for a certain element of unnaturalness that might be perceived across human-system interaction situations (alongside other challenges – e.g., Kanda & Miyashita 2016; Mavridis, 2015).

Path-related terms

This class of expressions encompasses prepositions like *across*, *through*, and *along*, all of which convey a sense of a path. This can be either dynamic, i.e., directly traversed as part of an ongoing motion process (as in *He walked through the room*), or static, which Talmy (2000) characterizes as 'fictive motion' (as in *The rug is spread out across the room*). In addition to the sense of extendedness associated with the concept of (actual or fictive) motion, each of these prepositions presupposes a certain dimensionality (Talmy, 2000). While *along* requires a one-dimensional path concept, *across* suggests two dimensions, and *through* appears to evoke a three dimensional scenario (as in *through the air*). These geometrical constraints considerably restrict frequency of usage in everyday language, and limit their feasibility in the smart environment scenario targeted here. Furthermore, path notions typically presuppose implicit understanding of essential path elements such as the start and goal position (Rojo & Valenzuela, 2003), where diverse path-related prepositions express only a subpart of the path: compare *from the kitchen / along the hallway / to the door.* Creating an actual path from an underspecified linguistic utterance that only represents partial path aspects calls for intricate inference procedures and probabilistic reasoning (Fasola & Matarić, 2013), and for intricate grounding processes in the case of human-robot interaction (Spranger et al., 2016). This limits their feasibility for facilitating joint reference in smart home scenarios.

Distance-related and other terms

Some prepositions express notions of distance, such as *near, far,* and *close.* The use of these prepositions requires concepts of distance relative to the experience of typical object-to-object relationships. Thus, an utterance like *The sink is close to the fridge* appears to suggest a spatial layout within a person's reach in a kitchen, whereas an utterance like *My house is close to the station* rather suggests that the station is either within walking distance or otherwise close enough to be feasible for easy travelling. The difficulty of predicting whether a context-dependent term of this kind would be considered appropriate in a particular situation has been noted time and again in the literature (e.g., Carlson & Covey, 2005), and this considerably hampers the feasibility of being used by automatic systems (Kelleher et al., 2006).

The linguistic repertory of spatial relational terms furthermore encompasses terms that do not fit into any one of the previously mentioned classes, such as *between* and *opposite*. Both of these require a complex spatial relationship between more than two objects, and this again constrains the feasibility of usage across situation contexts. Moreover, an automatic system would need to be equipped with the spatial knowledge required to determine when exactly an object can be described as being *between two others*: how flexible is the location relative to a line between the two reference objects, and how close to one of them could the target object be located? *Opposite* appears to come with even more intricate constraints, presupposing some notion of intrinsic sides facing each other across some boundary, obstacle, or distance (Bateman et al., 2010).

In spite of the constraints for using the spatial terms discussed so far, there is one main advantage that all of them share: They can be used independently of an observer's perspective. Thus, *the chair is at the table*, *the bread is on the plate, the sink is close to the fridge,* and *the plate is between the fork and the knife* could all be used without needing to consider where the speaker and listener are currently located, and from which angle they view the scene – or whether they view it at all. This advantage can be invaluable in a setting where perspectives are not clear. A smart environment, in particular, does not impose a particular point of view; instead, it is conceived of as ubiquitous, enabling interaction with the user at any location within the environment. This feature has considerable advantages, as

users will never need to take the system's perspective, and will not need to worry whether the system is able to interact at a particular location. Topological, path- and distance-related, and other spatial terms as just discussed all work very well with this kind of interaction setting, since they do not involve a particular perspective for their interpretation. However, each of them comes with clear and often experientially based restrictions of use, which means that they cannot be applied under all kinds of circumstances, and often require extensive world knowledge to be used successfully. This hampers their suitability for smart home interaction settings.

In the next section I will introduce another class of spatial terms, namely those denoting the various dimensions in surrounding space: *left, right, in front, behind, above,* and *below.* These terms offer ways of referring to spatial locations that are generally far more flexible across usage contexts, as they presuppose less complex spatial relationships, and depend less on experiential knowledge about specific object relationships and functions of object usage: in theory, any two objects could be described as being *left* or *right* of each other, as long as there are no intervening objects at the same level of granularity (Tenbrink, 2007). This set of terms is frequently referred to as *projective*, as the terms project directions from a specific point of view – i.e., they depend on the existence of a perspective. For this reason, they are regarded as cognitively more demanding than other spatial prepositions, posing further challenges for interpretation (Kelleher et al., 2006). Moreover, since the underlying perspective is typically implicit in language, this opens up various possibilities for misunderstandings based on underspecified reference. To explain these effects systematically, projective terms are typically discussed in terms of *spatial reference systems*. Apart from perspective-dependent projective terms, there is another class of reference systems that are less prominent in many people's minds and have rarely been considered in the context of automatic systems, but that may arguably pose interesting opportunities for smart-home scenarios: *absolute* reference systems, which are based on a ubiquitously accessible conceptual framework for assessing spatial locations, such as compass directions.

Spatial reference systems

Tidying up a wide range of earlier approaches with partially contradicting solutions to explain their scope, Levinson (2003) suggested a now widely used typology of reference frames encompassing *relative*, *intrinsic,* and *absolute* systems. The *absolute* category is typically distinguished by the use of a different set of terms such as those denoting compass directions; using an absolute reference system, a place can be located to another one as being *to the North* of it. In contrast, *intrinsic* and *relative* systems can be easily confused as they both draw on the same linguistic resources – projective terms. The main difference between the two systems consists of the existence of an external point of view (the *origin* in Levinson's terminology) in the case of *relative* reference: *the ball is in front of the table from Angela's point of view* relates the ball not only to the table but also explicitly refers to Angela as an origin of the underlying reference system. Similar observations hold if the speaker (*I*) or the listener (*you*) were used for this purpose. In contrast, an utterance like *the ball is in front of the chair* can be understood independently, without requiring an origin; here, the ball is related directly to the chair's front, and this would be true independent of where speaker and listener are located. In this sense, intrinsic reference systems require two positions in space (that of the target object, called *locatum,* and that of the object it is related to, called *relatum* by Levinson), whereas relative reference systems require three positions (locatum, relatum, and origin) (Tenbrink, 2011). However, both reference systems share the requirement that some kind of perspective or direction needs to be identified: either that of the relatum, or that of the origin. If none of these are readily and unambiguously available, reference using projective terms is not possible.

Relative and intrinsic reference systems have frequently been used in automatic systems for establishing spatial reference, addressing various associated challenges (cf. Schütte et al., 2017). For instance, Moratz & Tenbrink (2006) aimed to capture speakers' spontaneous strategies by computational modelling of overlapping acceptance areas for specific subsets of projective terms, using either relative or intrinsic interpretations. Ross & Kelleher (2014) addressed cases of reference frame ambiguity. Carrión & Sanfeliu (2014) targeted the automatic production of scene maps based on projective and topological descriptions, suggesting pragmatic solutions to the associated indeterminacies and ambiguities.

However, while informed guesses can often be made as to underlying perspectives and reference frames, ultimately the problem remains that projective terms are highly ambiguous and conceptually challenging, even for humans (Kelleher et al., 2006). Indeed, most people are intuitively aware of how easily *left* and *right* are confused in everyday situations, and of the possibility of misunderstanding the intended perspective (*my left or yours*?), in spite of extensive experience in spatial communication. Speakers frequently (and without warning) mix perspectives in description (Tversky et al., 1999), and they flexibly integrate multiple cues in their interpretation (Galati & Avraamides, 2013). A reliable, contextually adaptive integration of these intricate communicative processes in automatic systems thus remains a substantial future challenge.

In the light of these issues, it is worthwhile returning to the concept of *absolute* reference systems, which offer further opportunities for spatial reference. They do not require any kind of perspective at all, are more flexible than topological or distance terms, and do not come with any particular functional associations. Also, positioning objects in a global (absolute) framework corresponds more directly to the positioning techniques that automatic systems readily use, e.g. based on GPS (Hightower & Borriello, 2001). Nevertheless, absolute reference systems have rarely been considered in the context of smart home scenarios. This may be due to the fact that in the context of natural language use, discussions of absolute reference systems are typically associated with cardinal directions based on compass terms – and these are not readily available to potential users of smart homes. In many cultures (including some English-speaking regions) speakers rarely use compass terms at all, and in particular not in indoor environments where orientation happens within a smaller scale. This is different for some other cultures (Levinson, 2003), such as the Yucatec Maya. In a community where cardinal directions are essential for survival, speakers grow up knowing at all times how things are related to each other in a cardinal sense, independent of a specific viewpoint. These speakers can represent this knowledge through language or through gestures (Le Guen, 2011).

Although extremely versatile, this option is not available in most contexts where smart homes are likely to be built up in the near future. It may, however, be possible to use technological solutions to provide directional knowledge of the kind that would enable fully versatile absolute reference systems in smart home

scenarios. Some related applications are already under development. For instance, a magnetic belt is now on the market that constantly indicates the North direction through vibration. This kind of device is commercially available in various versions. Its effects on human spatial cognition have been researched in depth by the FeelSpace team (Nagel et al., 2005), suggesting that even without extensive previous experience in drawing on compass directions for navigation, humans are able to integrate this additional source of information in their everyday wayfinding processes.

While the main target application for the belt is navigation in outdoor environments, our current concerns open up novel opportunities for indoor (or small-scale) object reference, based on the knowledge of a constant direction. Considering our initial example, the ubiquitous assistive system may simply inform the user that *the pills are North of the cupboard*. Equipped with constant awareness of the North direction through the belt, the user should quickly be able to find the target object. References to other cardinal directions may be slightly more challenging, as they are not indicated by the belt. It is conceivable (though requires empirical exploration) that a certain amount of practice would lead to a similar status as with references to surrounding space, where the area *in front* is conceptually more prominent than any other area (Franklin, Henkel, & Zangas, 1995), and consequently *left* and *right* are more easily confused than *front* and *back*. Thus, the empirical prediction would be that although belt wearers would be most at ease with references to *North*, they would have little difficulty establishing quickly where *South* is, and would – with some practice and potential errors – also be able to interpret references to *East* and *West*. The belt transmits knowledge about the North direction on a constantly available physical basis, requiring no action and no conscious calculations; instead, it makes directional knowledge directly (subconsciously) available, comparable to the way humans always know where their front side is.

If wearing a belt is not feasible, other solutions may be available to serve similar purposes. A simple compass could be used to indicate North, requiring the user to adapt to the unconventional type of reference system via the use of an external device. The cognitive cost of this solution, which requires conscious calculation and the immediate availability of a compass, may be one reason why compass directions are rarely used in indoor scenarios in most English speaking cultures.

Somewhat simpler, another recent development is a device (https://beeline.co) that was designed for bicyclists to indicate the 'beeline' direction to their goal location through a simple arrow, independent of their current orientation. The device can indicate North just as well as any other direction, and thus provides stable directional information in a less physically immersive way than the belt, while being flexibly adaptable to various kinds of purposes – and less cumbersome than the use of a conventional compass. Indeed, it may be conceivable to use a similar device for pointing purposes, circumventing challenges related to language or other ways of representing direction.

To expand further on the idea of using stable directional systems, Levinson's notion of absolute reference systems is not restricted to the cardinal directions established by a compass. Brown & Levinson (1993), for instance, showed how prominent environmental features can constantly serve to establish reference based on an absolute system, such as *uphill* and *downhill* in Tzeltal. In that culture, it is common to refer to an object as being *uphill* from another object or a person, even in indoor scenarios. This is similar to compass-based cardinal directions, except for the fact that local speakers of Tzeltal would constantly be aware of the culturally central *uphill* and *downhill* directions, while speakers of English may or may not be aware where *North* is at any given time. Thus, arguably the main factor impeding the use of absolute reference systems in indoor scenarios is the lack of awareness of a relevant directional system – and this can be overcome by devices such as the magnetic belt or a prominently accessible compass system. In a smart home that is aligned with compass directions, the Northern wall could for instance be marked as such.

This idea relates to a suggestion by Pederson (2003), who argued that ad-hoc solutions based on perceptually available directions work in essentially the same way as fixed (culturally agreed) absolute reference systems. Pederson's example (2003:290) is, *The cat is towards the wall from the trashcan*. This would be equivalent to saying that *the cat is North of the trashcan*, (and in fact identical if the wall happens to be North of the trashcan). This suggests that absolute reference systems could be invented ad-hoc, offering flexible solutions for reference in the absence of awareness of a constant absolute system, or a perspective or other useful relationship such as those based on topological relations or distance.

Pederson's example seems limited, however, in that there would be more than one wall in a room, making this particular example highly ambiguous. Intuitively, the wall can still serve for reference if a particular wall is closer to the trashcan than the other walls. However, such intuitions would need to be implemented, and the assistance system may come to different conclusions than the user. If uniquely identifiable referents are used ad-hoc, this comes with the further complication of establishing unambiguous joint reference to both objects in question. Then, this kind of ad-hoc absolute reference system will have similar restrictions as the spatial term *between* discussed above: it will necessitate a specific and clearly discernible spatial relationship between two objects. Indeed, it may be more natural to say *between the wall and the trashcan* than *towards the wall from the trashcan.* However, if a certain wall was already clearly identified as the North wall and frequently used for reference, such complications would be avoided.

Spatial dialogue

To establish joint spatial reference with a ubiquitous system, it is necessary for the user and the system to draw on common ground, i.e. a certain amount of shared knowledge. Following insights from the previous section, the kinds of shared knowledge that are required to build up a meaningful spatial relationship include:

- the position of another object that the target object could be related to
- common sense notions of relative distance, functional relationships, and the like that affect the use of distance or topological terms
- a perspective in the case of using an intrinsic or relative reference system
- awareness of an absolute reference system such as compass directions, uphill/downhill, or a local one that enables similar reference.

While there have been many efforts to capture the notion of common ground in general (Clark, 1996) and in human-robot interaction settings (Chai et al., 2014), the computational management of situated spatial dialogue is still under-developed (Guadarrama et al., 2013; Skočaj et al., 2016) and requires creative solutions for reference handling (Mast, Falomir & Wolter, 2016), including attempts to incorporate the human's gaze in the system's interpretation procedure (Barz, Poller, & Sonntag, 2017), and strategies for handling errors (Schütte et al., 2017). One major challenge concerns the fundamental difference between human concepts represented by natural language, especially in the domain of space

(Bateman et al., 2010), and formal systems suited for computational purposes, e.g. spatial reasoning – even if based on qualitative rather than metric relations (Moratz & Ragni, 2008). Humans integrate a wide range of aspects of a spatial scene in order to interpret direction and distance concepts meaningfully, which is difficult to formalize in generic or context-adaptable terms (Rodrigues, Santos, & Lopes, 2016). A related issue concerns the many ways in which human communication is frequently underspecified (Van Deemter & Peters, 1996). While humans manage to fill in the many conceptual gaps relatively smoothly drawing on societally grounded communicative principles (Clark, 1996), automatic systems inevitably lack the world knowledge and specific action experience to do so in a reliable and intuitive manner, requiring intricate adaptation procedures (Chen, Yang, & Chen, 2016). These factors considerably complicate the establishment of common ground between humans and automatic systems.

In addition, communicative principles identified for human-human interaction do not necessarily transfer to human-system interaction settings, implying additional challenges for implementation. For instance, humans switch flexibly between spatial perspectives when interacting with another human, but not when interacting with an automatic system (Tenbrink et al., 2010). While humans would adopt the interaction partner's perspective primarily in the case of problems or discrepancies in spatial ability (Schober, 2009), automatic systems seem to be deemed generally incapable of adopting the user's perspective, and so users consistently use the system's perspective. It is an open empirical question how speakers would act in a situation where the system does not own a perspective at all, as in the case of a ubiquitous smart home assistance system.

One possibility may be to pro-actively suggest a perspective that the user can agree on, thus establishing common ground in this regard. Then, projective terms would be available for reference, using relative reference systems. Humans sometimes do this in spatial dialogue when perspectives are not currently shared or non-present spatial scenarios are discussed (Denis et al., 1999; Krause & Schiehlen, 2001). For instance, the system might say: *coming in to the kitchen, the pills are on the table to your left.* This is a fairly natural description that users should easily be able to handle. However, this strategy requires the identification of a suitable perspective that the system can detect and verbally describe so as to

enable unambiguous identification by the user – and this may be highly context dependent. Lacking the extensive experience and world knowledge that human speakers draw on when identifying suitable imaginary or currently accessible perspectives for reference, such a strategy may not be straightforward to implement in any sufficiently flexible, context-adaptive manner.

Traditionally, information-state based systems have tackled the common ground challenge through dialogue that incrementally establishes the current state of shared information (e.g., Traum & Larsson, 2003). In the context of a smart-home environment with everyday tasks such as searching for an object, users may prefer a direct answer; elaborate dialogues may not be feasible or desired. Instead, it may be more suitable to establish an agreed solution that can be re-used across many situations that involve spatial localization in the user's everyday life. Seen in this light, the use of an absolute reference system appears to be a feasible candidate, due to its flexibility and the lack of reliance on elaborate additional world knowledge. Once user and system have established the accessibility of a certain directional system as common ground, it can be added to the system's knowledge base, along with the set of relevant directions that can then be used flexibly. Such a generalized solution is not available with any other spatial term, since the specific requirements (distance, functional relationships, perspective, etc.) vary for each object-to-object relationship. In contrast, absolute reference systems remain stable and can be used to relate any two random objects to each other. Moreover, the availability of this solution would also enable the user to describe a spatial relationship independent of a perspective – solving the problem of how to adapt to the system's perspective when it does not actually have one.

Conclusion

In this paper, I have sketched the range of strategies and spatial terms available to establish reference in a smart home interaction scenario. Humans share sufficient common ground to be able to flexibly use heavily situation-dependent spatial expressions such as *left, on* or *near* in dialogue. The implementation of this kind of strategy in automatic systems is generally recognized as highly problematic, and it poses additional challenges where a ubiquitous system does not provide a perspective. However, the spatial limitations of a smart home environment offer an unconventional opportunity: to establish a unique internal *absolute* reference

system that can serve as a flexible and versatile basis for joint spatial reference across many everyday tasks and situations. Such a system could be based on compass directions, but other options are conceivable.

References

Barz, M., Poller, P., and Sonntag, D. 2017. Evaluating remote and head-worn eye trackers in multi-modal speech-based HRI. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 79-80. ACM.

Bateman, J., Hois, J., Ross, R.J., and Tenbrink, T. 2010. A linguistic ontology of space for natural language processing. *Artificial Intelligence 174*: 1027–1071.

Brown, P. and Levinson, S.C. 1993. "Uphill" and "Downhill" in Tzeltal. *Journal of Linguistic Anthropology, 3*: 46–74.

Carlson, L. and Covey, E.S. 2005. How far is near? Inferring distance from spatial descriptions. *Language and Cognitive Processes, Vol. 20*, No. 5, pp. 617 – 632.

Carrión, E.R. and Sanfeliu, A., 2014. Human-robot collaborative scene mapping from relational descriptions. In *ROBOT2013: First Iberian Robotics Conference,* pp. 331-346. Springer.

Chai, J.Y., She, L., Fang, R., Ottarson, S., Littley, C., Liu, C. and Hanson, K. 2014. Collaborative effort towards common ground in situated human-robot dialogue. In *Proceedings of the 2014 ACM/IEEE international conference on human-robot interaction*, pp. 33-40. ACM.

Chen, K., Yang, F. and Chen, X., 2016. Planning with task-oriented knowledge acquisition for a service robot. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pp. 812-818. AAAI Press.

Clark, Herbert H. 1996. *Using Language*. Cambridge: Cambridge University Press.

Coventry, K. R. and Garrod, S. C. 2004. *Saying, seeing and acting: The psychological semantics of spatial prepositions*. Hove and New York: Psychology Press.

Denis, M., Pazzaglia, F., Cornoldi, C., and Bertolo, L. 1999. Spatial discourse and navigation: an analysis of route directions in the city of Venice. *Applied Cognitive Psychology* 13:2, 145 – 174.

Fasola, J. and Matarić, M.J., 2013. Modeling dynamic spatial relations with global properties for natural language-based human-robot interaction. In *RO-MAN, 2013 IEEE,* pp. 453-460. IEEE.

Franklin, N., Henkel, L.A., and Zangas, T. 1995. Parsing surrounding space into regions. *Memory and Cognition, 23,* 397-407.

Freksa, C. 1992. Using Orientation Information for Qualitative Spatial Reasoning. In A. U. Frank, I. Campari, and U. Formentini (Eds.), *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, pp. 162-178. Berlin: Springer.

Galati, A., & Avraamides, M. N. 2013. Flexible spatial perspective-taking: Conversational partners weigh multiple cues in collaborative tasks. *Frontiers in Human Neuroscience*. DOI:10.3389/fnhum.2013.00618

Guadarrama, S., Riano, L., Golland, D., Go, D., Jia, Y., Klein, D., ... and Darrell, T. 2013. Grounding spatial relations for human-robot interaction. In *Intelligent Robots and Systems (IROS),* pp. 1640-1647.

Herskovits, A. 1986. *Language and spatial cognition.* Cambridge: Cambridge University Press.

Hightower, J. and Borriello, G. 2001. Location systems for ubiquitous computing. *Computer*, *34*(8), pp.57-66.

Kanda, T. and Miyashita, T., 2016. Communication for Social Robots. In *Cognitive Neuroscience Robotics A*, pp. 121-151. Springer.

Kelleher, J., Kruijff, G., & Costello, F. 2006. Proximity in context: an empirically grounded computational model of proximity for processing topological spatial expression. *Proceedings of COLING-ACL'06. Sydney, Australia. Association of Computational Linguistics.*

Krause, P., Reyle, U., and Schiehlen, M. 2001. Spatial inferences in a localization dialogue. In: M. Bras & L. Vieu (eds.), *Semantic and Pragmatic Issues in Discourse and Dialogue: Experimenting with CurrentDynamic Theories*. Amsterdam: Elsevier, pp. 183-216.

Le Guen, O. 2011. Speech and gesture in spatial language and cognition among the Yucatec Mayas. *Cognitive Science 35*:5, 905–938.

Levinson, Stephen C. 2003. *Space in Language and Cognition*. Cambridge University Press.

Mast, V., Falomir, Z., and Wolter, D. 2016. Probabilistic reference and grounding with PRAGR for dialogues with robots. *Journal of Experimental & Theoretical Artificial Intelligence*, 28(5), 889-911.

Mavridis, N., 2015. A review of verbal and non-verbal human–robot interactive communication. *Robotics and Autonomous Systems*, *63*, pp. 22-35.

Moratz, R. and Ragni, M. 2008. Qualitative spatial reasoning about relative point position. *Journal of Visual Languages and Computing 19*: 75-98.

Moratz, R. and Tenbrink, T. 2006. Spatial reference in linguistic human-robot interaction: Iterative, empirically supported development of a model of projective relations. *Spatial Cognition and Computation 6*:1, pp. 63-106.

Nagel, S., Carl, C., Kringe, T., Märtin, R., and König, P. 2005. Beyond sensory substitution – Learning with the sixth sense. *Journal of Neural Engineering 2*, 13-26.

Rodrigues, F.M.E., Santos, P.E. and Lopes, M., 2016. Communication of spatial expressions on multi-agent systems using the qualitative Ego-Sphere. In *Control and Automation (ICCA), 2016 12th IEEE International Conference on*, pp. 25-30. IEEE.

Rojo, A., and Valenzuela, J. 2003. Fictive Motion in English and Spanish. *International Journal of English Studies* 3(2): 123–49.

Ross, R.J. and Kelleher, J.D. 2014. Using the situational context to resolve frame of reference ambiguity in route descriptions. In *Proceedings of the Second Workshop on Action, Perception and Langauge (APL'2), Uppsala, Sweden*.

Schober, M.F. 2009. Spatial dialogue between partners with mismatched abilities. In K. Coventry, T. Tenbrink, and J. Bateman (Eds.), *Spatial Language and Dialogue*. Oxford University Press, pp. 23-39.

Schütte, N., Mac Namee, B. and Kelleher, J. 2017. Robot perception errors and human resolution strategies in situated human–robot dialogue. *Advanced Robotics*, pp.1-15.

Skočaj, D., Vrečko, A., Mahnič, M., Janíček, M., Kruijff, G. J. M., Hanheide, M., ... & Zillich, M. 2016. An integrated system for interactive continuous learning of categorical knowledge. *Journal of Experimental & Theoretical Artificial Intelligence, 28*(5), 823-848.

Spranger, M., Suchan, J., Bhatt, M. and Eppe, M., 2014. Grounding dynamic spatial relations for embodied (robot) interaction. In *Pacific Rim International Conference on Artificial Intelligence*, pp. 958-971. Springer.

Srimal, P. A. S., and Jayasekara, A. B. P. 2017. A multi-modal approach for enhancing object placement. In *National Conference on Technology and Management (NCTM),* pp. 17-22. IEEE.

Talmy, L. 2000. *Toward a Cognitive Semantics*, 2 vols. Cambridge, MA: MIT Press.

Tenbrink, T. 2007. Space, time, and the use of language: An investigation of relationships. Berlin: Mouton de Gruyter.

Tenbrink, T. 2011. Reference frames of space and time in language. *Journal of Pragmatics 43*:3, 704-722.

Tenbrink, T., Ross, R.J., Thomas, K.E., Dethlefs, N., and Andonova, E. 2010. Route instructions in map-based human-human and human-computer dialogue: a comparative analysis. *Journal of Visual Languages and Computing 21*:5, 292–309.

Traum, D. and Larsson, S. 2003. The information state based approach to dialogue management. In R. Smith and J. van Kuppevelt (Eds.), *Current and New Directions in Discourse and Dialogue*. Dordrecht: Kluwer, pp. 325-353.

Tversky, B., Lee, P., and Mainwaring, S. 1999. Why do speakers mix perspectives? *Spatial Cognition and Computation*, 1 (4), 399-412.

Van Deemter, K. and Peters, S. (Eds.), 1996. *Semantic Ambiguity and Underspecification*. CSLI, Stanford, CA.