Distribution is not enough: going Firther

Andy Lücking
University Paris-Diderot, Goethe University Frankfurt
luecking@em.uni-frankfurt.de

Robin Cooper
Gothenburg University
robin.cooper@ling.gu.se

Staffan Larsson
Gothenburg University
staffan.larsson@ling.gu.se

Jonathan Ginzburg
University Paris-Diderot
yonatan.ginzburg@univ-paris-diderot.fr

Abstract

Much work in contemporary computational semantics follows the *distributional hypothesis* (DH), which is understood as an approach to semantics according to which the meaning of a word is a function of its distribution over contexts which is represented as vectors (word embeddings) within a multi-dimensional semantic space. In practice, use is identified with occurrence in text corpora, though there are some efforts to use corpora containing multi-modal information.

In this paper we argue that the distributional hypothesis is *intrinsically* misguided as a *self-supporting basis* for semantics, as Firth was entirely aware. We mention philosophical arguments concerning the lack of normativity within DH data. Furthermore, we point out the shortcomings of DH as a model of learning, by discussing a variety of linguistic classes that cannot be learnt on a distributional basis, including indexicals, proper names, and wh-phrases. Instead of pursuing DH, we sketch an account of the problematic learning cases by integrating a rich, Firthian notion of dialogue context with interactive learning in signalling games backed by in probabilistic Type Theory with Records. We conclude that the success of the DH in computational semantics rests on a *post hoc* effect: DS presupposes a referential semantics on the basis of which utterances can be produced, comprehended and analysed in the first place.

1 Introduction

Much work in contemporary computational semantics follows the *distributional hypothesis* (DH), attributed to Harris (1954) and Firth (1957), which is understood as an approach to semantics according to which the meaning of a word is a function of its distribution over contexts which is represented as vectors (word embeddings) within a multi-dimensional semantic space. In practice, use is identified with occurrence in text corpora, though there are some efforts to use corpora containing multi-modal information (e.g Bruni et al., 2014). The appealing prospect of distributional semantics (DS) is to provide a *self-supporting semantic theory* according to which semantic representations can be bootstrapped from corpora in an entirely empirical fashion:

> Word space models constitute a purely descriptive approach to semantic modelling; it does not require any previous linguistic or semantic knowledge, and it only detects what is actually there in the data. (Sahlgren, 2005, p. 3)

In this paper we argue that the distributional hypothesis is intrinsically misguided as a self-supporting basis for semantics (for such a claim, see Baroni et al. (2014); also Baroni and Lenci (2010) argue in that direction), as—somewhat surprisingly—Firth was entirely aware. Note that our discussion points at reference and interaction as semantic building blocks, not to inferential and compositional properties which are often mentioned as motivating factors behind engaging in *Formal Distributional Semantics* (FDS; Boleda and Herbelot, 2016) or distributional probabilistic inference (Erk, 2016). In section 2 we begin with mentioning pertinent philosophical arguments concerning the lack of normativity within DH

data due to McGinn (1989) (and with reservations Kripke 1982). These argument show that collections of past uses do not allow to deduce a notion of veridicality, since the regularities within any collection of data do not project to semantic norms. Using supervised models (where target data or a so-called 'ground truth' are spelled out in advance), DS might try to implement normative knowledge nonetheless. But this move amounts to a circular approach: semantic knowledge is needed in order to apply distributional semantic methods in the first place. Thus, we observe a *bootstrapping problem* here. This problem is conceded within DS in general: 'The results suggest that, while distributional semantic vectors can be used "as-is" to capture generic word similarity, with some supervision it is also possible to extract other kinds of information from them [...]' (Gupta et al., 2015, p. 20). Obviously, the bootstrapping problem completely undermines the DH's claim to provide a semantic theory. However, work that aims at accounting for such 'higher order' phenomena in distributional terms but drawing on additional resources or annotations—like Herbelot and Vecchi (2015), who map distributional information onto quantified sentences, Gupta et al. (2015), who map distributions to knowledge base information, or Aina et al. (2018), who employ distributional semantics on character annotations within a specifically designed network model—puts its emphasis not on distributional semantic representation, but on how to *learn* these representations. We observe a further, *de facto* construal of DH here, a construal that draws on the notion of supervised learning, triggered by the procedures of data extraction or machine learning. Accordingly, in section 3 we discuss a variety of linguistic classes that cannot be learnt on the basis of the distributional hypothesis, including indexicals (Oshima-Takane et al., 1999), proper names (Herbelot, 2015), and wh-phrases. In section 5 we sketch an account of these problematic cases that integrates a rich notion of dialogue context (Ginzburg, 2012) with interactive learning in signalling games backed by a probabilistic Type Theory with Records (Cooper et al., 2015), which lays the foundations for spreading probabilities over situations.

This diagnosis of DS can be related to two different notions of semantics as a theory of meaning (cf. Lewis, 1970, p. 19). , namely semantic theory as a specification of the meanings of words and sentences (vector spaces), and a foundational theory of meaning identify the facts in virtue of which words and sentences do have the meaning they have (learning) We point at struggles of the DH with both of these respects here. We will even, for the most part, put aside one of the intrinsic issues confronting the DH—the *sentential denotation issue*: how to get distributional vectors to denote/represent/correspond to events or propositions or questions. Nor do we deal with compositionality. We deal primarily with lexical semantic issues (though not entirely, given the discussion of wh-phrases ...).

2 Philosophical Arguments against the distributional hypothesis in a nutshell

The normativity argument is as simple as powerful: since not every language use is a correct one, you want to partition uses into correct and incorrect ones. However, as ascertained by McGinn (1989, p. 160), 'this partitioning cannot be effected without employing a notion not definable simply from the notion of bare use'. That is, use falls short of capturing veridical normativity.[1]

Thus, the distributional hypothesis as pursued in computational semantics (textual or multimodal) fails to account for *semantic normativity*. Following teleological reasoning in philosophy of language (which takes off with Millikan, 1984) we suggest that meaning does not just reside in mere use, but is learned by trial and error under the pressure of coordination, as is partly simulated within evolutionary game theory. However, based on discussing three particular linguistic phenomena (Sec. 3), we further

---

[1]There is a second argument against usage-based semantics, which appeals to the problem of induction: Kripke (1982, Sec. 2) argues that use does not fix the extension of the reference relation by example of 'quus' and 'plus'. 'Quus' and 'plus' are two different arithmetic functions which have the same value up to a certain point but diverge thereafter. Having only observed uses of '+' up to the point of divergence, why should you refer by '+' to 'plus' and not to 'quus'? However, this argument is convincing only when fixed extensions are assumed, as is usually done in the closed models of formal semantics. A classifier-based semantics (Larsson, 2015), which is embraced by the authors, is compatible to open models and therefore takes Kripke's argument as further evidence that extensions indeed are not fixed.

argue with Firth that semantic games have to couched in an interaction based ontology (Sec.4).

3 Linguistic Evidence against the distributional hypothesis

To the more philosophical concerns (which alone point to the difficulties of taking DH in some form as the basis for semantic theory) we ground our case with three classes of linguistic expressions, all of which are pervasive in conversation and that seem to resist a distributional analysis, namely *indexicals*, *proper names*, and *wh-words*. Indexicals and proper names involve the issue of *semantic denseness*: they occur in too many places to have distinctive neighbours; this, arguably also afflicts wh-phrases, though they have distinctive syntactic positions in some languages.

An NLP adherent, who cares not about cognitive concerns, could argue for indexicals and wh-phrases that these need to be hard-wired in (and added as dimensions in the vector space) and that a distributional semantics deals with the lexical dynamics of a mature learner. This does not work so well for proper names, but once again one could claim that these are hard-wired in as independent dimensions (the entity libraries used by Aina et al. 2018 come close to this approach) and then new items learnt by analogy. But in any case, they would need to demonstrate that the emergent semantics could deal with the (dialogical) context dependence and perspectivity of all these expressions, which vector representations do not, at present at least, offer a solution to.

3.1 Indexicals

Examples of indexical expressions are the first and second person singular pronouns *I* and *you*. What does it take to acquire the corresponding semantic rules? Moyer et al. (2015, p. 2) are explicit in this regard: 'Thus, to achieve adult-like competence, the child must infer on the basis of the input, that *I* marks speaker, *you* marks addressee, and *s/he* marks a salient individual, usually a non-participant. To do so, she must pay attention to the discourses in which pronouns are used, specifically at each given moment, who is speaking, who is being addressed, and who is participating (or not) in the conversation.' The dialogue role-awareness is crucial in order for children to acquire the capacity of reversal necessary for the mastery of indexical pronouns, a difficulty ascertained by Oshima-Takane et al. (1999). That is, in case of indexical pronouns it is precisely not their co-occurrence pattern that give clues to their reference, but their relation to contextual indices. A learning game for pronouns that follows closely these psycholinguistic lines is designed in Sec. 5.1.

3.2 Proper names

With regard to proper names there is the notorious difficulty to separate named entity vectors from common noun vectors. This can only be achieved by employing additional processing on top of a distributional analysis like named entity recognition within a domain of uniquely named individuals (Herbelot, 2015; the need of additional preprocessing is also admitted by Baroni et al., 2014, p. 260, if one wants to distinguish proper names from common nouns at all). However, it is difficult to see how this unique description approach to proper names within distributional semantics (DS) can be applied to 'real-world' data where one and the same proper name word form (say, *John*) refers to many different individuals, intuitively, that individual that is jointly known by that name to the interlocutors. A very standard naming game that gives rise to polysemous names, but employs two vocabularies, is sketched in Sec. 5.2.

3.3 wh-words

With wh-words we reach a new difficulty that, in contrast to indexicals and proper names, they are not referential. Wh-words are used to from questions or initiate relative clauses. When used to form a question, a wh-word (like *which?* or *who?*) can address nearly any constituent or referent of the preceding text. In terms of distributions, wh-words therefore stand out due to contextual promiscuity.

What is lost is their indication of illocutionary force (question marking), which seems to be a key aspect.[2] Furthermore, wh-words can request implicit referents, that is, referents that lack a surface realisation.

An example being *A: I found an earring yesterday. B: Where?*, where B queries the location of the finding situation, which is not verbalised. Since DS operates on surface co-occurrence, any implicit referent (as any kind of elliptical construction) has to be regarded as theoretically problematic.[3]

4   Ontology

The phenomena discussed (indexicals, proper names, wh-words) rely on relations to constituents of the situation of utterance, contextual relations which Firth was fully aware of (cf. Firth, 1957, p. 5 f.). Firth distinguishes two main sets of relations: firstly, syntagmatic and paradigmatic relations; secondly situational relations, which include as a subset what he calls 'analytic relations':

> Analytic relations set up between parts of the text (words or parts of words, and indeed, any 'bits' or 'pieces'), and special constituents, items, objects, persons, or events within the situation. (Firth, 1957, p. 5, footnote omitted)

However, back in his time '[t]he technical language necessary for the description of contexts of situation is not developed' (Firth, 1957, p. 9).[4] This shortage has been remedied by contemporary semantic theory, in particular dialogue semantics. Therefore, we follow a more complete approach, as may have been envisaged by Firth, and employ a detailed context model (namely Ginzburg's (2012) KoS) in order to get a grip on indexicals, proper names, and *wh*-words. In this sense, we plead for 'going Firther' by incorporating the full range of contextual relations into semantic models. Very briefly, language use takes place in utterance situations which can be modelled in terms of *dialogue game boards*, that is, 'scoresheets' that keep track of participants, utterances and meanings (a very rudimentary example is provided in Sec.5.1 below—in fact, it is so rudimentary that it only hosts participants; utterances and meanings are recorded in the exchanges and utility functions of the accompanying game dynamics). Meanings are construed as *Austinian propositions* (going back to Austin, 1950 and adopted in situation semantics (Barwise and Perry, 1983)), pairs of situations and situations types (which will be crucial for accounting for indexicality in Sec.5.1). Similarly, following Ginzburg et al. (2014), we appeal to *Austinian questions* as pairs of situations and abstracts; their original motivation was a unified treatment of Boolean operations with propositions and an account of adjectival modification. Here, following Moradlou and Ginzburg (2014), we appeal to them as a means for conceptualizing how questions get acquired as a consequence of situationally grounded interaction. Thus, there are three kinds of semantic objects (*SemObj*), namely (objects in) situations (can also be understood as *frames*), Austinian propositions, and Austinian questions. Meanings then are mappings from utterance situations to SemObj and can be learned from aggregating experiences linking interactions in situations to SemObj. We sketch three scenarios thereof.

5   Learning lexical meanings

Using the ontology briefly exposited in the previous section, we sketch three learning scenarios for acquiring mastery of the meaning of the linguistic phenomena introduced in Sec. 3. Our strategy is to relate the ontological context model to game theory. Unlike Lazaridou et al. (2017), who aim at training agents in a game-theoretical setting, our motivation is theoretically driven by the semantic requirements to incorporate interaction and normativity. The respective learning task is operationalised in terms of agents' behaviour that, in the evolution of successful games, converge from random actions to utility-driven actions.

---

[2] wh-words also have embedded uses where the querying force is neutralized, an extra complication explicated by the semantic combinatorial mechanism.

[3] One might argue that distributions can still be obtained from annotated data where implicit referents are explicated. However, DS seems to be unable pull the required annotation off itself. So one needs a semantic analysis in order to get a distributional semantics to work in the first place—we identified this circularity as the bootstrapping problem in Sec. 1.

[4] On 'contexts of situation' see also Firth (1935, p. 64 ff.).

5.1 Pronoun games

Using singular pronouns properly depends on the speaker being a discourse participant or not, thus, it is intrinsically related to the utterance situation (this is the indexicality of first and second person singular pronoun). Regardless of which role a participant has in discourse, the first person pronoun refers to the speaker, the second person pronoun to the addressee and the third person one to some salient individual (bystander) different from speaker and addressee (assuming that *he* or *she* are used exophorically, that is, referring to an individual accessible in the utterance situation).

A child acquires the corresponding linguistic competence between one and three years (see the survey provided by Moyer et al., 2015, p. 2–3). This competence can be conceived as being acquired in 'pronoun games' that take place in different kinds of learning situations characterised by the discourse role and pronoun used. Let us assume a minimal example involving a tiny social network consisting of a child (girl) and its parents (say, traditional mother and father). The utterances that are produced in this situations are about someone being dirty, where someone is identified by one of three pronouns (for simplicity, we do not distinguish the gender of the third person pronoun (alternatively, one could assume a non-traditional family structure where just one third person is needed anyway)):

(1) a. 'I'm dirty.'

    b. 'You are dirty.'

    c. 'She/he is dirty.'

The declarative sentences from (1) give rise to Austinian propositions (cf. Sec. 4), namely that the situations thereby described are of the type claimed by the compositional derivation of the three word utterances. Let us further assume, that each participant has acquired a solid competence in judging an individual as being of type *dirty* in a situation (we assume for concreteness that this is accomplished by means of Bayesian learning in probabilistic TTR according to Cooper et al. (2015) exploiting perceptual classifications (Larsson, 2015).

Thus, the blueprint of the pronoun game is a situation of the following form, which is more or less the format of an utterance type in KoS (Ginzburg, 2012):

(2)
$$
s : \begin{bmatrix}
\text{phon} & : & \langle\ \alpha\ \text{am/are/is dirty}\ \rangle \\[2pt]
\text{dgb-params} & : & \begin{bmatrix}
\text{spkr} & : & \text{Ind} \\
\text{addr} & : & \text{Ind} \\
\text{bystander} & : & \text{Ind} \\
\text{s0} & : & \text{Rec} \\
\text{c1} & : & \text{addressing(spkr,addr)} \\
\text{c2} & : & \text{disengaged}(\{\text{spkr,addr}\}, \text{bystander})
\end{bmatrix} \\[2pt]
\text{cont} & : & \begin{bmatrix}
\text{sit} & = & \text{dgb-params.s0} : \text{Rec} \\
\text{sit-type} & = & \begin{bmatrix}
\text{x} & = & \begin{array}{l}\text{s.dgb-params.spkr} \vee \\ \text{s.dgb-params.addr} \vee \\ \text{s.dgb-params.bystander}\end{array} & : & \text{Ind} \\
\text{c}_h & : & \text{dirty(x)}
\end{bmatrix} : \text{RecType}
\end{bmatrix}
\end{bmatrix}
$$

Given the schema in (2), the actual learning situation for a pronoun game is determined by randomly instantiating the discourse roles 'spkr', 'addr' and 'bystander' with a member of the social network (child, mother, father), and by assigning label x a value from the social network (i.e., s.dgb-params.spkr, s.dgb-params.addr, s.dgb-params.bystander), fixing who is dirty and is referred to.

Thus, given the modelling of Firthian 'contexts of situation' by means of modern dialogue semantics (Cooper, 2019; Ginzburg, 2012, and Sec. 4), all ingredients being there for engaging in a *Lewis game*

(Lewis, 1969), following standard game theory in semantics and pragmatics (Steels, 1997; Jäger, 2012). We construct such a game along the lines of Mühlenbernd and Franke (2012) by means of the following parameters:

- Social signals are exchanged between a *sender* (spkr) and a *receiver* (addr).

- Exchanges take place in a states $t_i \in T$, where each t_i is defined by assigning individuals to spkr, addr, and bystander ($3! = 6$, by combinatorics) and fixing x (3 possibilities). So the sample space T has $6 \times 3 = 18$ states to choose from. Since all have an equal chance to be chosen, the probability p for each $t_i \in T$ is $p(t_i) = \frac{1}{18} \approx 0.055$.

- The signals exchanged are messages $m \in \{I, you, she\}$, $p(m = I) = p(m = you) = p(m = she) = \frac{1}{3}$.

- Messages are responded to by an action $a \in \{I, you, she\}$, which, when effected by mother or father, amounts to parental feedback (amounting to answering 'right' or 'wrong').

- Players want to communicate successfully, so each participant has a utility function $u(t_i, a_j) = \begin{cases} 1 & \text{if } t_i = a_j \\ 0 & \text{otherwise} \end{cases}$

Current (and future) exchanges are informed by a history of successful, past interactions—in terms of a sender's belief about a receiver $B_r(a \mid m)$ and a receiver's belief about a sender $B(t \mid m)$—, stored in expected utility functions *EU* for both sender and receiver ((3) and (4) closely follow the functions defined by Mühlenbernd and Franke (2012)):

(3)   a. $EU_s(m \mid t) = \sum_{a \in A} B_r(a \mid m) \times u(t, a)$

      b. $EU_r(a \mid m) = \sum_{t \in T} B_s(t \mid m) \times u(t, a)$

A *best response strategy* σ for the sender (given a situation type t) and ρ for the receiver (given sender's message m) is to maximise their expected utilities. A standard method in order to compute this strategies is the $\arg\max$ function:

(4)   a. $\sigma(m \mid t) = \begin{cases} \frac{1}{\arg\max_m EU_s(m\mid t)} & \text{if } t \in \arg\max_m EU_s(m \mid t) \\ 0 & \text{else} \end{cases}$

      b. $\rho(a \mid m) = \begin{cases} \frac{1}{\arg\max_a EU_r(a\mid m)} & \text{if } a \in \arg\max_a EU_r(a \mid m) \\ 0 & \text{else} \end{cases}$

In order to validate this model it is not even necessary to run a simulation study. The pronoun game employs a fixed vocabulary without homonyms and synonyms; each participant has equal chances to be assigned to the discourse roles; and the utility function acts as an amplifying function: such naming games are known to lead to lexical convergence (cf. De Vylder and Tuyls, 2006).[5] Since in our case the parents outnumber the child and therefore have greater impact on the expected utility functions, we can also predict which equilibrium will eventually be reached and characterise it with the lexical extracts in (5):[6]

---

[5] Readers might think of the pronoun games in terms of Pólya urns: each speaker has a urn associated with each type $t_i \in T$ that contains the n copies of the lexical elements of choice for denoting t_i (likewise for recipients). After a successful exchange took place, the number of copies of the successful lexical item will be increased by k, while the unsuccessful ones will be decreased by m ($n, i, k \in \mathbb{N}$ being parameters of the game). Thus, on the next turn, chances for converging on a lexical item increase already by number, and so on.

[6] Unless, of course, the parents entertain a private language where the meanings of the singular pronouns are interchanged in some way. But there will be lexical convergence in such a case, too.

(5) $\begin{bmatrix} \text{phon} : \langle\, \text{i}\, \rangle \\ \text{dgb-params} : \begin{bmatrix} \text{spkr} : \text{Ind} \\ \text{addr} : \text{Ind} \\ \text{bystander} : \text{Ind} \end{bmatrix} \\ \text{cont} : \begin{bmatrix} \text{x=dgb.spkr} : \text{Ind} \end{bmatrix} \end{bmatrix}$, $\begin{bmatrix} \text{phon} : \langle\, \text{you}\, \rangle \\ \text{dgb-params} : \begin{bmatrix} \text{spkr} : \text{Ind} \\ \text{addr} : \text{Ind} \\ \text{bystander} : \text{Ind} \end{bmatrix} \\ \text{cont} : \begin{bmatrix} \text{x=dgb.addr} : \text{Ind} \end{bmatrix} \end{bmatrix}$, $\begin{bmatrix} \text{phon} : \langle\, \text{she/he}\, \rangle \\ \text{dgb-params} : \begin{bmatrix} \text{spkr} : \text{Ind} \\ \text{addr} : \text{Ind} \\ \text{bystander} : \text{Ind} \end{bmatrix} \\ \text{cont} : \begin{bmatrix} \text{x=dgb.bystander} : \text{Ind} \end{bmatrix} \end{bmatrix}$

That is, the pronoun *I* refers to the speaker of the utterance situation, *you* to the addressee and *she/he* to the bystander, irrespective whether the discourse roles are occupied by father, mother, or child.

In the implementation of Aina et al. (2018), which is related to our proposal, pronouns are learned from semantically annotated dialogue of character references (obtained from data of the TV series *Friends* of Chen and Choi 2016) by means of an LSTM with an *entity library*. The input data looks as follows, where numbers are IDs for characters:

JOEY (183): '...see Ross (335), because I (183) think you (335) love her (306).'

The authors assume that 'the LSTM can learn to simply forward the speaker embedding unchanged in the case of pronoun *I*' (Aina et al., 2018, p. 68). But simply mapping *I* to whatever is left to the colon still misses perspectivity, which is the crucial issue about first and second person pronouns, cf. Moyer et al. (2015).

5.2 Proper names

At first glance, proper names seem to be the easiest part of speech to learn. Indeed, it is individual constants (predicate logic's formal device for representing proper names) that are learned in the most basic naming game scenarios (Lücking and Mehler, 2012). However, as briefly discussed in Sec. 3.2, proper names are homonymous expressions: one and the same name (understood as a word form) can apply to different individuals. In order to reflect this in a naming game, the assumption of a one-to-one correspondence between a word and an object has to be given up in favour of set-valued predicate constants. However, lifting proper names to polysemous terms clearly misses their point of being identificational expressions. Some semantic theories suggest to reconcile this conflict by analysing proper names as a hybrid of referential and descriptive expressions, where a proper name poses a presuppositional constraint on the individual referred to, which is (in a successful exchange) individuated by being mapped to a description (for such a semantic analysis see Cooper, 2019, Chap. 4). In order for agents of a naming game to learn such 'presuppositional names', we propose a two-stage game resting on two disjoint vocabularies: At the first stage, agents acquire possibly polysemous predicates by interchanging symbols from the first vocabulary. At the second stage, agents expands the $1 : m$-relations learned in the first level by means the second vocabulary. The second stage aims at capturing the descriptive content of proper names. We conjecture that the second stage could also be implemented in terms of grounding types by means of Bayesian learning (Cooper et al., 2015). In effect, agents acquire a second set of naming conventions which in most cases disambiguate the first set.[7] Granted, this is just the first step into the build-up of a long-term memory concerning named individuals (cf. Cooper 2019, Chap. 4; see also Weston et al. 2014).

Dialogically speaking, in this two-stage set-up agents acquire the equipment that partly corresponds to exchanges like the following: A: Sam was injured. B: Which one? A: The fireman. What is missing is the learning of wh-question (and answering them), to which we turn now.

5.3 Wh-words

Moradlou and Ginzburg (2014) sketch an account of how question understanding emerges. They identify a three-stage process:

---

[7]There may still be ambiguity left, but this also happens in real life and indicates that still further descriptive information have to be gathered in some cases.

1. Salient Object Identification games: a question is asked prompting for an appropriate descriptor of an object presented to the child.

2. Erotetically Plausible question games (EPQ) games: a question is asked in a situation where an obvious question arises.

3. Situational Description (SD) games: a question is asked about properties of objects observable in the situation.

Stages 1 and 2 habituate the child to associate wh-words with the need to consider possible resolutions of questions. In stage 1 the hypothesis space ranges over properties of objects. In stage 2 the hypothesis space is broadened to other aspects of the situation. The required SemObj become available due to the semantic ontology employed (cf. Sec. 4).

Wh-word competence ultimately consists in the identification of the queried domain and recognition that a question is being posed. Our basic conjecture is

(6) Conjecture: the sequence SOI–EPQ–SD leads to the emergence of wh-word competence.

We discuss here the first stage of this sequence, namely SOI games. An SOI game can be implemented on top of a two-stage naming game as sketched in Sec. 5.2. When an interaction in terms of a word from the first vocabulary is ambiguous, the receiver has two response actions: Firstly, she can randomly pick an object which is associated with the word as learned so far. Secondly, she can request a word from the sender's second vocabulary for the sender's referent. It is likely that the additional naming leads to an unequivocal identification of the object talked about. The utility function rewards successful identification. Thus, in all evolutionary simulations where the second vocabulary serves better than chance, requesting it (i.e., asking a wh-question) turns out to be the preferred strategy.

6   Conclusions and Future Work

We started the paper by reviewing a classic philosophical argument relating to normativity against basing semantics on sampling language use—a position embraced by distributional semantics. This argument is complemented by pointing out three classes of linguistic expressions which we suggested are intrinsically non-distributional. Making the case that meanings have to be learned in interactive language acquisition, both philosophical arguments as well as linguistic phenomena involving indexicals, proper names and wh-questions, suggest intrinsic problems for a semantics driven by the DH. Somewhat ironically, a construction plan appropriate for accounting for the linguistic phenomena under consideration can be found in the works of Firth in terms of 'contexts of situation'—although DS folklore hands Firth down exclusively as a founding figure of the distributional hypothesis. Accordingly, drawing on a semantic ontology of utterance situations, semantic objects and interactions, learning scenarios for indexicals, proper names and wh-questions are sketched that take Firth's notion of context seriously. While a semantics based on the DH suffers from shortcomings with respect to providing an adequate account of semantic learning, we conceptualise such an account by arguing for combining an ontology grounded in interaction with evolutionary game theory. To spell out the approach sketched here, future work needs to include running evolutionary simulation studies for those settings for which there is as yet no proof that they lead to lexical convergence.

We also identified a bootstrapping problem for DS that contributes to preventing it from being a self-supporting basis for semantics. But how does this assessment conform to the obvious success of DS in computational semantics? The reason, we think, is that a foundational semantics gives DS a piggyback: once there is independently justified meaning, usage regularities can be observed *post hoc*. We further hypothesise that this foundational semantics rests on interaction in context.

Acknowledgements

References

Aina, L., C. Silberer, W. M. Sorodoc, Ionut-Teodor, and G. Boleda (2018). AMORE-UPF at SemEval-2018 task 4: BiLSTM with entity library. In *Proceedings of the 12th International Workshop on Semantic Evaluation*, SemEval-2018, Stroudsburg (PA), pp. 65–69. ACL.

Austin, J. L. (1950). Truth. In *Proceedings of the Aristotelian Society. Supplementary*, Volume xxiv, pp. 111–128. Reprinted in John L. Austin: *Philosophical Papers*. 2. ed. Oxford: Clarendon Press, 1970.

Baroni, M., R. Bernardi, and R. Zamparelli (2014). Frege in space: A program of compositional distributional semantics. *LiLT (Linguistic Issues in Language Technology) 9*.

Baroni, M. and A. Lenci (2010). Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics 36*(4), 673–721.

Barwise, J. and J. Perry (1983). *Situations and Attitudes*. Stanford: CSLI Publications.

Boleda, G. and A. Herbelot (2016). Formal distributional semantics: Introduction to the special issue. *Computational Linguistics 42*(4), 619–635.

Bruni, E., N. K. Tran, and M. Baroni (2014). Multimodal distributional semantics. *Journal of Artificial Intelligence Research 49*, 1–47.

Chen, H. Y.-H. and J. D. Choi (2016). Character identification on multiparty conversation: Identifying mentions of characters in tv shows. In *Proceedings of the SIGDIAL 2016 Conference*, pp. 90–100.

Cooper, R. (2019). Type theory and language: From perception to linguistic communication. Book Draft.

Cooper, R., S. Dobnik, S. Larsson, and S. Lappin (2015). Probabilistic type theory and natural language semantics. *Linguistic Issues in Language Technology – LiLT 10*(4), 1–43.

De Vylder, B. and K. Tuyls (2006). How to reach linguistic consensus: A proof of convergence for the naming game. *Journal of Theoretical Biology 242*(4), 818–831.

Erk, K. (2016). What do you know about an alligator when you know the company it keeps? *Semantics and Pragmatics 9*(17), 1–63.

Firth, J. R. (1935). The technique of semantics. *Transactions of the Philological Society 34*(1), 36–73.

Firth, J. R. (1957). A synopsis of linguistic theory, 1930–1955. In *Studies in Linguistic Analysis* (Repr. 1962 ed.)., pp. 1–32. Oxford: Blackwell.

Ginzburg, J. (2012). *The Interactive Stance: Meaning for Conversation*. Oxford: Oxford University Press.

Ginzburg, J., R. Cooper, and T. Fernando (2014, April). Propositions, questions, and adjectives: a rich type theoretic approach. In *Proceedings of the EACL 2014 Workshop on Type Theory and Natural Language Semantics (TTNLS)*, Gothenburg, Sweden, pp. 89–96. Association for Computational Linguistics.

Gupta, A., G. Boleda, M. Baroni, and S. Padó (2015). Distributional vectors encode referential attributes. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, EMNLP'15, pp. 12–21.

Harris, Z. (1954). Distributional structure. *Word 10*(23), 146–162.

Herbelot, A. (2015). Mr Darcy and Mr Toad, gentlemen: distributional names and their kinds. In *Proceedings of the 11th International Conference on Computational Semantics*, pp. 151–161.

Herbelot, A. and E. M. Vecchi (2015). Building a shared world: Mapping distributional to model-theoretic semantic spaces. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 22–32.

Jäger, G. (2012). Game theory in semantics and pragmatics. In C. Maienborn, K. von Heusinger, and P. Portner (Eds.), *Semantics: an International Handbook of Natural Language Meaning*, Volume 3 of *Handbücher zur Sprach- und Kommunikationswissenschaft*, Chapter 4, pp. 2487–2516. De Gruyter Mouton.

Kripke, S. A. (1982). *Wittgenstein on Rules and Private Language: An Elementary Exposition*. Cambridge, MA: Harvard University Press.

Larsson, S. (2015). Formal semantics for perceptual classification. *Journal of Logic and Computation 25*(2), 335–369.

Lazaridou, A., A. Peysakhovich, and B. Marco (2017). In *Proceedings of the International Conference on Learning Representations*, ICLR 2017.

Lewis, D. (1969). *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

Lewis, D. (1970). General semantics. *Synthese 22*(1/2), 18–67.

Lücking, A. and A. Mehler (2012). What's the scope of the Naming Game? Constraints on semantic categorization. In *The Evolution of Language. Proceedings of the 9th International Conference*, Evolang IX, Kyoto, Japan, pp. 196–203.

McGinn, C. (1989). *Mental Content*. Oxford and New York: Basil Blackwell.

Millikan, R. G. (1984). *Language, Thought and Other Biological Categories*. Cambridge, MA: MIT Press.

Moradlou, S. and J. Ginzburg (2014). Learning to understand questions. In P. Muller and V. Rieser (Eds.), *Proceedings of SemDial 2014 (DialWatt)*, Heriot Watt University.

Moyer, M., K. Harrigan, V. Hacquard, and J. Lidz (2015). 2-year-olds' comprehension of personal pronouns. In E. Grillo, K. Jepson, and M. LaMendola (Eds.), *Boston University Conference on Language Development 39: Online Proceedings Supplement*, pp. 1–11. Boston Universtiy.

Mühlenbernd, R. and M. Franke (2012). Signaling conventions: Who learns what where and when in a social network? In *The Evolution of Language. Proceedings of the 9th International Conference*, Evolang IX, Kyoto, Japan, pp. 242–249.

Oshima-Takane, Y., Y. Takane, and T. R. Shultz (1999). The learning of first and second person pronouns in English: network models and analysis. *Journal of Child Language 26*(3), 545–575.

Sahlgren, M. (2005). An introduction to random indexing. `http://eprints.sics.se/221/1/RI_intro.pdf`.

Steels, L. (1997). The synthetic modeling of language origins. *Evolution of Communication 1*(1), 1–34.

Weston, J., S. Chopra, and A. Bordes (2014). Memory networks. `https://arxiv.org/abs/1410.3916`.