

---

# Formalizing the Dynamics of Semantic Systems in Dialogue

STAFFAN LARSSON

---

## 1 Abstract

Semantic change happens both in the long term (over years and decades) and in the very short term (in a single dialogue)<sup>1</sup>. An exploratory study of a Map Task dialogue indicates that dialogue participants coordinate on an ad-hoc vocabulary and associated concepts (meanings) to enable information exchange, and that ad-hoc vocabularies can be cobbled together from a heterogeneous mix of “micro-vocabularies” borrowed from various other (a priori unrelated) domains.

To account for these observations, I outline an abstract and general formal account of how the meanings of natural language expressions (typically, words) can change as a result of their use in dialogue. The account has two parts, one semantic and one pragmatic. In the semantics part, a basic framework for formalising updates to semantic systems in dialogue is sketched. In the pragmatics part, semantic updates are related to interactional processes of semantic coordination in dialogue, such as feedback, negotiation and accommodation.

## 2 Introduction

Semantic change happens both in the long term (over years and decades) and in the very short term (in a single dialogue). In addition, semantic change can occur either globally in a language such as Swedish or English, or more locally in specific institution- or activity-specific sub-languages, in social communities, or even between pairs (dyads) of speakers.

As an example, take the discussion in [Clark and Clark, 1979] of “porch” as a verb, as e.g. in “he porched the newspaper”. This is an example of an innovative use of an existing word, whose meaning thus changes (in this case, by being extended), provided that the innovation is understood and accepted by the other dialogue participants on concrete occasions of use. This may either be unproblematic (if the other speakers understand and accept the innovative use) or may involve

---

<sup>1</sup>The work presented here was funded by the Faculty of Arts, University of Gothenburg, and by the project “Semantic Coordination in Dialogue” (The Bank of Sweden Tercentenary Foundation project P2007-0717).

explicit negotiation of the appropriateness of using “porch” in this way. The resulting semantic change may either be limited to a specific dialogue (thus amounting to a “local convention”); it may spread over a community and eventually become part of the language as it is represented in dictionaries; or it may become part of a more limited subcultural/activity-specific “sub-language”, e.g. the jargon used by U.S. mailmen.

As a further example, an exploratory study of a Map Task dialogue (below) indicates that dialogue participants (henceforth, DP’s) coordinate on an ad-hoc vocabulary and associated concepts (meanings) to enable information exchange, and that ad-hoc vocabularies can be cobbled together from a heterogeneous mix of “micro-vocabularies” borrowed from various other (a priori unrelated) domains.

To account for these observations, I outline a formal account of how the meanings of words<sup>2</sup> can change as a result of their use in dialogue. This account has two main parts: (1) a semantic part in the form of an abstract account of *semantic plasticity* intended to capture formally how meanings can be modified (updated), and (2) a pragmatic account of *semantic negotiation* in dialogue, intended to capture how meanings are negotiated in dialogue and how such negotiations are related to meaning updates.

### 3 Background

This section presents the theoretical and empirical background and motivation for the theory proposed in this paper.

Formal semantics [Portner and Partee, 2002] gives precise analyses of meaning and compositionality, and formal pragmatics gives detailed accounts of how the “common ground” [Clark and Brennan, 1990] is dynamically updated in dialogue [Traum and Larsson, 2003]. However, these research areas typically assume that meanings are static and unchanging during the course of a dialogue, and have not paid much attention to the dynamics of meaning.

In the last two decades, several psycholinguists have begun turning their attention to lexical and semantic coordination in dialogue [Clark and Wilkes-Gibbs, 1986; Garrod and Anderson, 1987; Pickering and Garrod, 2004] and the dynamics of language use in limited groups [Brennan and Clark, 1996; Healey, 1997], and have provided evidence that linguistic conventions can emerge during the course of a dialogue or sequence of dialogues. Preliminary informal and semi-formal analyses have been presented, but the psycholinguistic tradition is not primarily concerned with developing comprehensive formal accounts of their findings. And as observed earlier, these are phenomena that traditional formal semantics and pragmatics are ill-equipped to deal with, because of their static view of meaning.

<sup>2</sup>Although this account is intended to cover not only words but also other expressions (phrases, syntactic categories, and other linguistic elements), “word” is frequently used instead of “linguistic expression”.

Some relevant computational work is being done [Steels and Belpaeme, 2005; Briscoe, 2002] on emergent vocabularies and category formation, proving that it is possible to simulate the dynamics of language and meaning in computers. However, this work has not yet addressed the wide variety of strategies for meaning negotiation available to human speakers, nor attempted to provide a comprehensive formal semantics (independent of, e.g., specific learning algorithms).

Thus, a general, detailed, empirically substantiated, and formal account of how linguistic meaning is negotiated and coordinated on a dialogue game micro-level (word-by-word, utterance-by-utterance) is still lacking. The work reported here is a first step towards such an account.

#### 4 Vocabulary in a Map Task dialogue

In the Map Task corpus<sup>3</sup>, a GIVER explains a route, provided on the giver's map, to a FOLLOWER who has a similar (but slightly different) map but with no route marked. A map contains landmarks portrayed as labelled line drawings. In a route-giving task like that recorded in the Map Task corpus, expressions referring to landmarks, compass directions etc. can be a priori expected as a kind of "prototype" devices for talking about maps. A typical utterance may look as follows<sup>4</sup>:

- (1) GIVER: right **a camera shop**, right, head due **south** ... from that just ...  
**down** for about **twelve centimetres**, have you got **a parked van** at the bottom ?

Here, we may note two expressions expressing direction ("south", "down"), one expressing a distance ("twelve centimetres") and two referring to landmarks ("a camera shop", "a parked van"). A further example:

- (2) GIVER: go round the left hand side of the camera shop ... in between **the edge of the page** and the camera shop.

Whereas the previous expressions were completely expected given the general direction-giving task, the reference to an absolute position using "the edge of the page" is perhaps less expected. Clearly, this is a consequence of the DP's talking about a (paper) map rather than e.g. about some actual terrain.

- (3) GIVER: so you're ... you're going diagonally sort of north ... northeast ...  
it's not it's it's a sort of **two o'clock** almost **three o'clock** ... from the allotments ... over

<sup>3</sup><http://www.hcrc.ed.ac.uk/maptask/maptask-description.html>

<sup>4</sup>The following excerpts are taken from Map Task dialogue q4nc4, available at the Map Task web site.

Here, we have GIVER referring to map directions using the expressions “two o’clock” and “three o’clock”. This is most likely an everyday variant of the practice of English-speaking pilots of using “o’clock” for directions<sup>5</sup>. Let’s look at a final excerpt:

- (4) GIVER: right, you go ... down the side of the camera shop right for about twelve centimetres ... and do a sort of **a ’u’ shape** ... for and **the bottom of the ’u’ shape** should be about three centimetres long, right do you know what i’m meaning  
 ... GIVER: you’ve worked it out already , eh we’re doing **a ’u’ shape** round the parked van but it’s a sort of three cent– see if you imagine a ’u’ right ... **the stems of the ’u’ the ... vertical bits** are sort of three centimetres between

First, a trajectory is referred to using the expression “a ’u’ shape”. This trajectory is then (or so it appears) reified as an imagined ’u’-shape on the map, now acting more akin to a landmark with a concrete (if invisible) shape, size and even component parts (“the ... vertical bits”; “the stems of the ’u’ ”).

## 5 Micro-vocabularies used in Map Task dialogue

Based on the above excerpts (and others from the same dialogue), we can provide a very tentative inventory of referring expressions used by GIVER and FOLLOWER in the Map Task dialogue. DP’s refer to distances, absolute and relative locations, directions, and trajectories. Below, we list the sub-types of expressions used for each basic class.

- distances on page, in centimetres (“about twelve centimetres”)
- absolute locations
  - landmarks (“the camera shop”)
  - page edges (“the edge of the page”; “at the bottom”; “the far right-hand side”)
  - typography on page (“the words ’yacht club’”)
  - (imagined) letter shapes (“the bottom of the ’u’ shape”; “the stems of the ’u’ the ... vertical bits”)
- relative locations
  - relative to landmark (“left hand side of (landmark)”)

---

<sup>5</sup>Note the use of a hedging “sort of” before “two o’clock”, which seems to indicate that the speaker is slightly unsure as to whether the following expression is quite appropriate. A similar observation is made by [Brennan, 2000] (p. 11): “[h]edges seem to be one way of marking that a referring expression is provisional.”

- relative to sheet of paper (“the other side of the page”)
- directions
  - compass directions (“head due south”)
  - left, right, up, down, diagonally, etc.
  - clock directions (“sort of two o’clock”)
- trajectories
  - imagined/drawn lines (“a straight line up the ...”)
  - letter shapes as trajectories (“do sort of a ‘u’ shape”)

Now, how can we account for this diversity in the range of linguistic expressions used in a simple direction-giving dialogue?

## 6 Interleaving resource registers

In this section, a basic terminology is proposed, intended to form a basis for a formal account of what we see happening in dialogues such as the one quoted above.

### 6.1 Perspectives

In the Map Task dialogue, the DP’s need to coordinate on a way of talking about the map. What the above excerpts show is that there are several ways of talking about a map; this is also shown in the Maze Game experiments [Garrod and Anderson, 1987; Healey, 1997] where DP’s alternate between an abstract “coordinate system” perspective on a maze (“Go to the fourth row down and the second from the right”; “Six three”), and more concrete perspectives involving e.g. corridors (“Go forward, then turn left at the junction”) or shapes (“the bit sticking out on the right”). A way of talking about  $X$  can be said to involve *taking a perspective*<sup>6</sup> on  $X$  and selecting a vocabulary associated with that perspective. Taking a perspective  $P$  on subject matter  $X$  in dialogue involves an analogue - “talking about  $X$  as  $P$ ” - e.g. talking about directions on a map as clock arms. Different perspectives have different advantages and disadvantages; for example, an abstract perspective is compact but error-prone; a clock perspective on directions may e.g. enable shorter utterances. One plausible reason for interleaving and switching several perspectives and associated vocabularies thus seems to be that it increases the efficiency of communication.

---

<sup>6</sup>This terminology follows [Clark, 1997]. [Garrod and Anderson, 1987] and [Healey, 1997] instead talk about adopting “description types”.

## 6.2 Resource and ad-hoc registers

A language can be regarded as consisting of a multitude of activity-specific “language games” involving *registers*<sup>7</sup>. This term will be used here to denote *activity-specific semantic systems* (“micro-languages”), each consisting minimally of a set of linguistic signs, i.e., linguistic expressions and associated concepts (meanings)<sup>8</sup>. In dialogue, registers may be used as *resources* which can be borrowed or appropriated into a new activity and adapted to the domain at hand. Putting it differently, an *ad-hoc register* is assembled to be able to talk about some subject matter from one or more perspectives. In the map-task dialogue, several different resource registers are introduced and accepted<sup>9</sup>. Often, both introduction and acceptance are implicit, but sometimes verbal signals (including feedback) are used to manage semantic coordination. For example, one could imagine the expression “sort of” being used (perhaps only indirectly, by expressing doubt that the adjacent term is appropriate) to signal introduction of new register.

As mentioned, in the Map Task dialogue we find some resource registers that can be regarded as “standard” or “default” ways of talking about maps, whereas others are more unexpected. First, the standard map register subsumes (1) a *landmarks* register provided to DP’s as pictures and text on map, (2) a *compass directions* register, and (3) a (*metric*) *distance* register. The non-standard parts of the ad-hoc register include:

- *clock* register: map directions as clock hands “two o’clock” etc.
- *sheet-of-paper* register perspective: map as a sheet of paper edges of page distances on page relations between pages (e.g. “opposing page”)
- *letter shape* register perspective: Viewing map as a piece of paper where letter shapes can be drawn letter shapes (“a ’u’ shape”) parts of letter shapes (“stems”)

## 6.3 Appropriating and interleaving registers

To describe the dynamics of registers in the above dialogue, we can say that the clock, sheet-of-paper and letter-shape registers are appropriated into the map task activity, where it is *interleaved* with landmark, compass direction, and metric distance registers to form an ad-hoc register<sup>10</sup>. This involves adapting the meanings associated with resource register vocabularies to the current situation.

<sup>7</sup>The term “register” is borrowed from [Halliday, 1978]. In general, a register (in the linguistic sense) is a subset of a language used for a particular purpose or in a particular social setting.

<sup>8</sup>A *compositional* register will more generally contain *mappings* between expressions and meanings.

<sup>9</sup>Often, several resource registers are used in a single phrase, as e.g. in “in between the edge of the page and the camera shop”.

<sup>10</sup>This “interleaving strategy” can be compared with the “switching strategies” evident in maze game experiments (Healey, Garrod), where speakers switch between perspectives (description types). Presumably, both interleaving and switching are possible.

## 7 Meaning potentials

To describe how linguistic expressions can be interactively (in dialogue) appropriated into a new activity, we need an account of semantics which (1) allows several activity-specific meanings for a single expression, and (2) allows open and dynamic meanings which can be modified as a consequence of language use.

The received view in formal semantics [Kaplan, 1979] assumes that there are abstract and context-independent “literal” meanings (utterance-type meaning; Kaplan’s “character”) which can be regarded formally as functions from context to content; on each occasion of use, the context determines a specific content (utterance-token meaning). Abstract meanings are assumed to be static and are not affected by language use in specific contexts. Traditional formal semantics is thus ill-equipped to deal with semantic coordination, because of its static view of meaning.

In the present account, we explore the possibility that the idea of “meaning potentials” may offer a more dynamic view of meaning. The term originates from “dialogical” approaches to meaning [Recanati, 2003]. On the “dialogical” view, language is essentially dynamic; meaning is negotiated, extended, modified both in concrete situations and historically. Interaction and context are essential for describing language, and there is a general focus on the context-dependent nature of meaning. Linguistic expressions have meaning potentials, which are not a fixed and static set of semantic features, but a dynamic potential which can give rise to different situated interpretations. Different contexts exploit different parts of the meaning potential of a word.

In the account of meaning potential presented in these pages, the term *semantic plasticity* refers specifically to the dynamic aspect of meaning potentials. This will be central to our account of how activity-specific abstract<sup>11</sup> meanings are updated and gradually change as a consequence of use<sup>12</sup>.

## 8 Towards a formalisation of semantic plasticity

To describe in more detail how DP’s coordinate on registers (e.g. when adapting a resource register to a new domain), we need a dynamic account of meanings and registers allowing incremental modifications (updates) to semantic systems. We also need a description of possible dialogue strategies for register coordination. Describing this process *formally* requires formalising the dynamics of registers and meaning potentials, and the dialogue protocols involved in negotiating semantic systems. In this section, I will take some initial steps towards this goal by sketching a formal account of semantic plasticity.

<sup>11</sup>I use “abstract meaning” to refer to utterance-type meanings, either activity-specific or activity-independent.

<sup>12</sup>Note that “dynamic semantics” [Groenendijk and Stokhof, 1988] is not dynamic in this sense, as it follows traditional formal semantics in assuming a static mapping between words and meanings.

To keep the theoretical framework general, it intentionally leaves open the issue of semantic representation (formulae of first-order logic, semantic features, semantic vectors, neural nets, etc.) and the kinds of learning mechanisms used in linguistic coordination. The main purpose of the formal semantic component of the theory is (at the current stage at least) not to champion one semantic theory over another. Instead, I want to provide an abstract and general semantic framework which can be connected to the pragmatic account of semantic negotiation to describe the interactive conditioning of usage patterns.

The proposal here is to regard the meaning of a linguistic expression or word to depend on previous uses of that word<sup>13</sup>. This makes it possible to model how meanings change as a result of using language in dialogue. The basic idea is that speakers have internalised (potentially complex) dispositions, or *usage patterns*, governing the use of specific words. These dispositions depend, among other things, on observations of previous situations where the word in question has been used, and on specific generalisations over these situations.

Semantic plasticity is described in terms of updates to individual usage patterns associated with words triggered by observations of their use in dialogue. When a usage pattern for a word  $c$  is sufficiently coordinated (shared) within a community, we can talk about  $c$  as having a meaning potential. Meaning potential is thus construed as emerging from processes of interactive coordination of usage patterns. By modelling plasticity of usage patterns of individuals, we thus indirectly model semantic plasticity in a linguistic community.

That a usage pattern connected to an expression is “sufficiently coordinated” in a community means, roughly, that speakers and hearers are able to use that expression to exchange information sufficiently to enable them to achieve their shared and private goals. For example, in the Map Task dialogues an expression is sufficiently coordinated when DP’s are able to make use of it in successfully carrying out the route-giving tasks assigned to them.

### 8.1 Usage sets and usage patterns

To get a handle on semantic plasticity, a *usage-set*<sup>14</sup>  $S_c^A$  will be posited for each language user  $A$  and word  $c$ , containing all situations<sup>15</sup> where  $A$  has observed a use (token) of  $c$ . Formally, this can be written as follows:

<sup>13</sup>This idea is not new. Its origins can be traced back to the idea that “meaning is use” [Wittgenstein, 1953].

<sup>14</sup>An alternative term is *situation-collocation*.

<sup>15</sup>It is important to point out that the notion of “situation” I am using here is an abstract one; the reason is that I want to keep the framework general. In more concrete instantiations of this abstract framework, the notion of a situation will be specified based on the activity in which an agent acts and the requirements on the agent in this activity, as well as the representations and sensory-motor machinery of the agent. As a simple example, in the work of [Steels and Belpaeme, 2005] the situation is limited to a colour sample, perceived by a robot through a camera and processed into a representations of colours in the form of three real-valued number.

- (5)  $S_c^A = \{s \mid A \text{ has observed a use of } c \text{ in situation } s\}$ .

The usage set should be regarded merely as an abstract theoretical entity. Again, similar suggestions have been made before:

Semantic potential is defined in terms of a set of source-situations. A source-situation is a situation that a speaker has learned to associate with a term because during learning the speaker has seen the term legitimately applied to that situation. [Bezuidenhout, 2002]

I assume that  $A$  generalises over  $S_c^A$  to produce a usage pattern (or usage disposition)  $[c]^A$ . In cognitive terms one can think of the usage pattern as the “memory trace” of observed uses of  $c$  by  $A$ .

### 8.2 Situated meanings and interpretations

On each occasion of use of  $c$  in situation  $s$ ,  $c$  has a specific situated (utterance-token) meaning, formally written as  $[c]_s$ , which derives partly from the shared abstract utterance-type meaning (meaning potential)  $[c]$  and partly from  $s$ . The subjective counterpart of a situated meaning is a *situated interpretation*, written as  $[c]_s^A$  for an agent  $A$ ; this is the interpretation that  $A$  makes of  $c$  in  $s$  based on  $A$ 's usage pattern  $[c]^A$ . A situated meaning  $[c]_s$  arises in a situation when DP's make sufficiently similar situated interpretations of  $c$  in  $s$ .

### 8.3 Appropriate and non-appropriate uses

I will assume that new uses of a word  $c$  can be classified as *appropriate* or *inappropriate* given an existing usage pattern<sup>16</sup> for  $c$ <sup>17</sup>. The formal notation which will be used to express that a use of  $c$  in situation  $s$  is appropriate with regard to  $A$ 's usage pattern for  $c$  is  $[c]^A \vdash s$ . Correspondingly,  $[c]^A \not\vdash s$  means that  $s$  is not an appropriate situation in which to use  $c$  given  $[c]^A$ <sup>18</sup>.

On the whole, if a token of  $c$  uttered in a situation  $s$  is consistent with  $[c]^A$ ,  $A$  is likely to understand  $c$  and to judge  $s$  to be an appropriate situation of use of  $c$ .

<sup>16</sup>It may be thought that appropriateness should be defined in terms of collective meaning potentials rather than individual usage patterns, to make sense of talk of “incorrect use of words.” However, I believe that such talk is better regarded as one of many strategies for explicit negotiation of meanings, which always occurs in concrete situations and between individual DP's with their respective usage patterns. A theoretical notion of correct or incorrect use of words (independent of individual usage patterns) runs into several problems, such as defining how many DP's must share a usage pattern in order for it to be deemed “correct.” This does not mean we cannot make sense of talk of incorrect and correct use of words; it only means that regard such notions primarily as devices in negotiations of shared meanings.

<sup>17</sup>In general, appropriateness is not necessarily a Boolean property, but rather a matter of degree. This is a simplification in the current formalisation.

<sup>18</sup>The exact method of deciding whether a new token is appropriate or not will depend on the specific kinds of representations, learning algorithms, and measures of similarity that are assumed (or, in an artificial agent, implemented).

However, it is important to leave open the possibility that a DP may not understand, or understand but reject, a token of  $c$  even if this token of  $c$  in the current situation is appropriate with respect to  $A$ 's usage pattern for  $c$ . Similarly, a DP may choose to use a word in a situation where she judges it inappropriate given previous uses; This can be referred to as a *creative use*, in contrast to *conservative uses* which are appropriate given previous uses.

#### 8.4 Usage-pattern updates

It follows from the definition of  $[c]^A$  that whenever  $A$  observes or performs a use of  $c$ ,  $S_c^A$  will be extended, and so the usage pattern  $[c]^A$  may change. This is a *usage pattern update*. *Prima facie*, there are many different possible kinds of ways that a usage pattern may be modified, depending on assumptions regarding semantic representation.

Usage-pattern updates can be distinguished according to several dimensions. An initial rough distinction can be made between *reinforcements* and *revisions*.

If a use of  $c$  in situation  $s$  is consistent with  $A$ 's usage pattern for  $c$ , i.e.,  $c$  is appropriate in  $s$  ( $[c]^A \vdash s$ ), there is no drastic change; the previous disposition is reinforced by extending  $[c]^A$  with  $A$ 's situated interpretation of  $c$  in  $s$ ,  $[c]_s^A$ . This will be written formally as  $[c]^A \circ_{=} [c]_s^A$ . However, if the current use of  $c$  is not consistent with usage disposition ( $[c]^A \not\vdash s$ ), there will be a relatively drastic revision of the disposition (formally,  $[c]^A \circ_{*} [c]_s^A$ ).

We may also want to be able to distinguish between updates based on positive (successful communication) and negative evidence (failed communication). This gives us four different possible update operations:

- $\circ_{=}^{+}$ : reinforce with positive evidence
- $\circ_{=}^{-}$ : reinforce with negative evidence
- $\circ_{*}^{+}$ : revise with positive evidence
- $\circ_{*}^{-}$ : revise with negative evidence

#### 8.5 Situation-types and structured meaning potentials

To account for how registers can be appropriated (borrowed) from one activity (e.g. telling the time) to another (e.g. direction-giving) we need a formalisation which allows new meanings of existing words to be created as a result of observed novel (at least subjectively) language use. Meaning potentials, in addition to being dynamic, can also be *structured*, and thus allow for different contexts to exploit different meaning potential *components*.

I will use *situation-type* as a general term for contexts, activities, institutions etc. where words take on specific meanings. A register, or “micro-language”, is the lexicon used in a situation-type, pairing the words used (vocabulary) with

meanings (what can be talked about; ontologies; coordinated usage patterns) in the situation-type<sup>19</sup>

In general, a situation-type may be associated with several registers (corresponding to different perspectives on the situation-type), each providing a mapping from a vocabulary to (abstract) meanings specific to the situation-type. Conversely, the meaning potential for a word is often structured into several situation-type-specific components.

Above, it has been established that  $[c]^A$  is agent  $A$ 's usage pattern for word  $c$ , and that  $[c]_s^A$  is the interpretation that agent  $A$  makes of  $c$  in  $s$ ; this interpretation is a function of  $s$  and  $[c]^A$ . This notation will now be extended with  $[c]_\alpha^A$  - an agent  $A$ 's situation-type-specific usage pattern component for  $c$  in situation-type  $\alpha$ . In general, any aspect of the utterance situation-type may activate usage pattern components. A structured meaning potential exists in a linguistic community with coordinated structured usage patterns. A component of structured meaning potential for  $c$  in situation-type  $\alpha$  is written as  $[c]_\alpha$ .

As a simple example inspired by the Map Task dialogue above, the meaning potential ["two o'clock"] can be described as structured into

- ["two o'clock"]<sub>clock</sub>, where *clock* stands for an activity type involving telling the time; this meaning potential component can be paraphrased "02:00 AM or PM"
- ["two o'clock"]<sub>dir-giv</sub>, where  $\alpha$  has been assigned a situation type index corresponding to direction-giving activities; this meaning potential component is paraphraseable as "east-northeast direction"

### 8.6 Interpretation and update involving structured usage patterns

A token  $c_s$  of a word  $c$  in situation  $s$  is interpreted by  $B$  as  $[c]_s^B$ . If  $[c]_s^B$  is a complex usage pattern, some component of  $[c]_s^B$  must be selected as the abstract meaning to be used for contextual interpretation. Now, assume that situation  $s$  is classified by  $B$  as being of situation-type  $\alpha$ . This triggers a component of  $[c]_s^B$  - the *activated usage pattern component*  $[c]_\alpha^B$ .

In this case,  $[c]_\alpha^B$  is a likely candidate for which part of  $[c]$  gets updated. (If  $B$  is not able to find a relevant usage pattern component,  $B$  may create a new ad-hoc component, which can be updated during the dialogue. This pattern may or may not be retained afterwards; it may be assimilated into some existing component of  $[c]$ , or the start of a new usage pattern component.)

Let's take an example. Assume the usage pattern ["two o'clock"] is structured into ["two o'clock"]<sub>clock</sub> and ["two o'clock"]<sub>dir-giv</sub>, as above. Now assume we get the following utterance:

<sup>19</sup>This terminology builds on (and modifies slightly) that of [Halliday, 1978].

(6) GIVER: “sort of two o’clock”

Because the activity is direction-giving, FOLLOWER activates the usage pattern component  $[\text{“two o’clock”}]_{dir-giv}^{follower}$ . FOLLOWER then instantiates the component  $[\text{“two o’clock”}]_{dir-giv}^{follower}$  to arrive at a situated interpretation  $[\text{“two o’clock”}]_s^{follower}$  (roughly, a 60 degree angle on FOLLOWER’s map). Insofar as  $[\text{“two o’clock”}]_{dir-giv}^{follower} \vdash s$ , we get a reinforcing update  $[\text{“two o’clock”}]_{dir-giv}^{follower} \circ_{\pm}^+ [\text{“two o’clock”}]_s^{follower}$ .

## 9 Semantic coordination

This section sketches a framework for modelling *negotiation of meaning in dialogue*, i.e. the social processes (dialogue games) involved in the explicit and implicit negotiation of meaning in dialogue, and their relation to the cognitive processes (semantic updates).

After discussing the basic devices available to speakers for conducting semantic negotiation, I will give examples of how the theory sketched above can be used to analyse short dialogue excerpts in terms of semantic updates. As yet, the theory does not include a taxonomy of dialogue moves involved in semantic negotiation, and therefore the analysis does not include dialogue moves; instead, utterances are analysed directly in terms of their associated semantic updates. Coming up with a general taxonomy of such moves and their associated updates is a major future research goal.

I assume (provisionally) three basic devices available to dialogue participants for negotiating (and, typically, achieving coordination of) linguistic resources: feedback, explicit negotiation, and accommodation. “Negotiation” is used here in a weak sense of “interactive achievement of coordination”.

### 9.1 Feedback

Feedback [Allwood, 1995; Clark, 1996] involves signals indicating perception, understanding, and acceptance of utterances in dialogue, as well as failure to perceive or understand; clarification requests; and rejections. It is well known that feedback governs that coordination of the dialogue gameboard (“informational coordination”); however, it also guides coordination of language use (“language coordination”).

*Corrective* feedback [Clark, 2003] is common in adult-child interaction. Below is an example; *A* is the child, *B* the adult, and as part of the common ground there is a topical object in the situation *s* visible to both *A* and *B*. Assume also that *A* is not familiar with the word “panda”<sup>20</sup>.

(7) A: Nice bear

<sup>20</sup>This is a made-up example; similar examples from real adult-child interactions can be found in [Clark, 2003].

B: Yes, it's a nice panda

$A$ 's initial situated understanding of "bear" will here involve various aspects of the topical object that appear relevant to  $A$ , e.g. that it looks like (has certain physical characteristics in common with) previous animals that people have called "bear" in the presence of  $A$ .

In Example 7,  $B$  rejects this use of "bear" by providing negative feedback in the form of a correction (however,  $B$  also gives positive feedback accepting the assertion that the focused object (animal) "is nice". After  $B$ 's utterance,  $A$  will produce a situated interpretation of "panda",  $[\text{"panda"}]_s^A$ . According to the principle that "speakers assume that any difference in form signals a difference in meaning" [Clark, 2003],  $A$  will try to detect a difference between bears and pandas (such as colour); presumably this will be part of  $A$ 's situated interpretation of "panda" which is used to revise (or in this case, create). In our theory, the semantic updates resulting from this dialogue are:

- $[\text{"bear"}]_s^A \circ_*^- [\text{"bear"}]_s^A$
- $A$  creates a new usage pattern  $[\text{"panda"}]_s^A$
- $[\text{"panda"}]_s^A \circ_*^+ [\text{"panda"}]_s^A$

The first update revises  $[\text{"bear"}]_s^A$  with  $s$  as negative evidence; this should make it less likely that  $A$  will use "bear" in similar situations in the future. It is possible that  $A$ 's updates will reflect an inference that a panda is a kind of bear. In any case, what  $A$  has learnt is that it is more appropriate to use "panda" than to use "bear" in situations similar to  $s$ . As for the second update, perhaps this new usage pattern will be based on  $[\text{"bear"}]_s^A$ ; this amounts to an assumption that pandas are similar to bears. The third update adds  $s$  as positive evidence of a use of "panda".

## 9.2 Explicit negotiation

Explicit negotiation is the overt meta-linguistic negotiation of the proper usage of words, including e.g. cases where explicit verbal or ostensive definitions are proposed (and possibly discussed). Although semantic negotiation typically has the goal of coordinating language use, it may in general be both antagonistic and cooperative.

Steels and Belpaeme [2005] investigate the effect of social linguistic interaction on learning and coordination of colour categories. Robot agents play a language game of referring to and pointing to colour samples. The language system of an individual agent is modelled as a set of categories in the form of neural nets that respond to sensory data from colour samples, and a lexicon connecting words to categories.

This is clearly a case of semantic plasticity and semantic negotiation, as categories are updated as a result of language use. Semantic negotiation here takes the form of explicit and cooperative negotiation. There is also an asymmetry with respect to the roles within each game; one agent is speaker and the other is the hearer. The interaction follows a predefined “guessing game” script; essentially, a language game of guessing and ostensive definition. The situation, as perceived by the agents, is a set of objects (colour samples), where one is the *topic* object.

Below is a possible interaction between two agents playing the guessing game, and the corresponding updates in terms of the model presented in this paper.

The context is  $O$  a set of object (colour samples),  $O = \{o_1, \dots, o_N\}$  where one object  $o_t$  is the focus object. The speaker ( $A$ ) knows which object is the focus object but the hearer ( $B$ ) does not. The goal of the game is for  $B$  to correctly identify  $o_t$  from  $O$  based on the interaction with  $A$ .  $A$  will first utter the word (here: “wabaku”) that  $A$  associates with  $o_t$ , i.e. the word associated to the semantic category (in the form of a neural network) that uniquely discriminates  $o_t$ . In terms of the theory outlined in these pages, I identify the neural network with the usage pattern, which for the speaker is [“wabaku”]<sup>A</sup>.

(8) A: wabaku

$B$  now looks up “wabaku” in its own lexicon and finds an associated semantic category [“wabaku”]<sup>B</sup>.  $B$  then produces a situated interpretation [“wabaku”]<sub>s</sub><sup>B</sup>, namely (given the architecture of agents in this experiment)  $B$ ’s sensory impression of an object ( $o_g$ ) that  $B$  (in this example) is able to uniquely pick out from the context using the semantic category (neural network) [“wabaku”]<sup>B</sup>.

(9) B: (points at object  $o_g$ )

$A$  then gives feedback to check whether this interpretation is correct. Unfortunately,  $B$  has made a wrong guess, which leads to negative feedback and a correction:

(10) A: (feedback indicating rejection of B’s answer)

A: (points to topic object  $o_t$ )

$B$  will now produce a new situated interpretation [“wabaku”]<sub>s</sub><sup>B</sup>, to be (the sensory impression from) object  $o_t$ . As a consequence of this game of semantic negotiation,  $B$  will then revise its meaning-pattern (neural network) [“wabaku”]<sup>B</sup> by adjusting it to better match this new situated meaning; in our theory<sup>21</sup>.

$$\bullet \frac{[\text{“wabaku”}]^B \circ_*^+ [\text{“wabaku”}]_s^B}{[\text{“wabaku”}]^B \circ_*^- [\text{“wabaku”}]_s^B}$$

<sup>21</sup>It seems that negative evidence is not used in this case; in our notation this would amount to the update [“wabaku”]<sup>B</sup>  $\circ_*^-$  [“wabaku”]<sub>s</sub><sup>B</sup>.

Note that accommodation is not an option in this game. One reason for this is that the context is not rich enough to allow the hearer to infer the intended referent in problematic situations without consulting the speaker.

### 9.3 Accommodation

In linguistics, *accommodation* can be taken generally to refer to adaptations to the behaviour of other DP's. For example, one may adapt to the presuppositions of an utterance of “The King of France is bald” by modifying the dialogue gameboard to include the existence of a king of France. For the purposes of the current account, the notion of accommodation will be extended beyond the dialogue gameboard, to include changes in the language system.

Upon hearing a word  $c$  in new situation  $s$ ,  $A$ 's reaction (the kind of feedback  $A$  gives) partly depends on  $[c]$ , but crucially,  $A$ 's behaviour is not *determined* by  $[c]$ . This means that  $A$  can understand, and may choose to accept, uses of  $c$  that deviate from  $[c]$ ; this is a case of accommodation. For understanding to be possible here, it is necessary that the hearer has access to contextual cues in the utterance situation that makes it possible to make a reasonable guess at the meaning of this token of  $c$ . Similarly,  $A$ 's own future uses of  $c$  partly depend on  $[c]$  but not determined by  $[c]$ . Thus,  $A$  can use  $c$  in ways that deviate from  $[c]$ . Also,  $A$  can correctly understand words that she does not use herself<sup>22</sup>.

Assuming a semantics which allows new uses of a word  $c$  to be classified as appropriate or inappropriate given an existing usage pattern for  $c$ , we can construct a simple table of the possible outcomes of a limited class of meaning negotiation games. This class is limited to cases where either a usage is understood and accepted (successful uses), or is rejected or not understood (unsuccessful uses). In the latter case, explicit negotiation of the meaning and proper usage of  $c$  may well occur; this table shows a subset of the cases where this does not happen, i.e. the negative feedback (regarding understanding or acceptance) is the last word on the matter (for the time being).

For each word  $c$  used in an utterance  $u$ , the addressee (here,  $B$ ) in a dialogue is (usually) expected to react if he thinks  $c$  was inappropriately used. If  $B$  is able to construct a situated interpretation  $[c]_s^B$  (which may involve more or less effort) but finds this use inappropriate ( $[c]^B \not\prec s$ ), this may be due to a mismatch between  $s$  (as perceived by  $B$ ) and  $[c]^B$ .  $B$  may now reject this use of  $c$  explicitly using negative feedback, or quietly alter  $[c]^B$  ( $[c]^B \circ_*^+ [c]_s^B$ ) so that this use of  $c$  can be counted as appropriate after all. The latter process we may call usage accommodation, or meaning accommodation.

As an example, take the (made-up) dialogue (11) where  $A$  uses “porch” in a situation  $s$  to mean approximately “successfully throw onto a porch”. We assume that  $B$  is not familiar with this use (and thus  $[”porch”]^B \not\prec s$ ) but that this is a

<sup>22</sup>However, it seems unlikely that anyone could correctly use a word that she does not understand.

	$[c]^A \vdash s$	$[c]^A \not\vdash s$
$[c]^B \vdash s$	default case $[c]^A \circ_{\pm}^+ [c]_s^A$ $[c]^B \circ_{\pm}^+ [c]_s^B$	unnoticed creative $[c]^A \circ_{*}^+ [c]_s^A$ $[c]^B \circ_{\pm}^+ [c]_s^B$
$[c]^B \not\vdash s$	accommodated conservative $[c]^A \circ_{\pm}^+ [c]_s^A$ $[c]^B \circ_{*}^+ [c]_s^B$	accommodated creative $[c]^A \circ_{*}^+ [c]_s^A$ $[c]^B \circ_{*}^+ [c]_s^B$

Table 1. Possible outcomes of speaker  $A$  successfully using  $c$ , addressed to hearer  $B$  in situation  $s$ .

common usage for  $A$  (and thus  $[^{\text{porch}}]^A \vdash s$ ).

- (11) A: On my paper round this morning I porched all the papers without getting off my bike!  
 B: Congratulations.

In this case,  $B$  is able to create a situated interpretation  $[^{\text{porch}}]^A_s$  by using contextual cues, world knowledge, commonsense background, and the ability to reason by analogy and metaphor. Also,  $B$  chooses to accept this novel use of “porch” and provides feedback displaying understanding and acceptance (and possibly other attitudes as well). This thus counts as a case of accommodated conservative use in Table 1, and as a consequence (or so our theory predicts) the corresponding updates are made. Other cases are of course also possible. For example, if  $B$  had not been able to understand  $A$ ’s use of “porch”, or if  $B$  had understood it but not accepted it, explicit negotiation would be expected.

#### 9.4 Accommodation of complex meaning potentials

Let us now have a look at a further example of semantic coordination, this time based on the Map Task dialogue, where meaning accommodation leads to updates to complex usage patterns.

Assume, as before, that we get the following utterance in a Map Task dialogue in a situation  $s$ :

- (12) GIVER: “sort of two o’clock”

In contrast to the example in Section 8.6, we now assume that FOLLOWER is *not* familiar with the “direction-giving” use of “two o’clock”. More precisely,  $[^{\text{two}}$

o'clock"]<sup>follower</sup> only contains ["two o'clock"]<sub>clock</sub><sup>follower</sup>, so ["two o'clock"]<sup>follower</sup>  $\not\vdash s$ .

By analogical reasoning using contextual features, FOLLOWER is nevertheless able to correctly understand A's utterance and arrives at a contextual interpretation ["two o'clock"]<sub>s</sub><sup>follower</sup>.

Now, since ["two o'clock"]<sup>follower</sup>  $\not\vdash s$ , FOLLOWER will need to revise ["two o'clock"]<sup>follower</sup> by creating a new activity-specific usage pattern component ["two o'clock"]<sub>dir-giv</sub><sup>follower</sup>. We get an overall update ["two o'clock"]<sup>follower</sup>  $\circ_{\pm}$  ["two o'clock"]<sub>s</sub><sup>follower</sup> which can be decomposed as two updates. First, creation of ["two o'clock"]<sub>dir-giv</sub><sup>follower</sup>, and then updating with the situated interpretation: ["two o'clock"]<sub>dir-giv</sub><sup>follower</sup>  $\circ_{\pm}$  ["two o'clock"]<sub>s</sub><sup>follower</sup>. After this update, we have ["two o'clock"]<sup>follower</sup>  $\vdash s$ , i.e., the novel (for FOLLOWER) use of "two o'clock" by GIVER has been accommodated.

## 10 Kinds of coordination in dialogue

On our view, two kinds of coordination happen in everyday human-human dialogue. *Informational coordination* has successfully been studied using the concepts of dialogue games and updates to a shared dialogue gameboard. One of the goals of the research presented here is to extend this approach to describing *language coordination* (and more specifically, semantic coordination) in terms of the dynamics of updates to language systems.

The framework sketched here aims at describing all kinds of semantic coordination. In the "two o'clock" example given above, coordination is essentially a matter of mapping an expression ("two o'clock") to a pre-existing meaning (denoted in the compass directions register as "east-northeast"). For this kind of coordination, some version of traditional formal semantics may suffice, provided it is extended with a dynamic mapping between linguistic expressions and their meanings.

However, in other cases the dynamics go beyond word-meaning mappings. Specifically, to account for cases where an expression is used to denote a *new* concept, such as "the 'u'-shape" above, we need to describe the dynamics of *concept creation*. Similarly, existing concepts may be affected by their use in dialogue, e.g., by subtly modifying values of usage-governing conceptual features by small increments. For example, in [Steels and Belpaeme, 2005], concepts are represented as neural nets which are updated by small adjustments to network weights, according to a standard back-propagation algorithm.

These dynamics, which I refer to as *concept-level* dynamics, are an important motivation for the introduction of meaning potentials. They are also our main reason for believing that traditional formal semantics will not suffice to account for semantic coordination.

To deal with concept-level dynamics in a general way, one will probably need to

keep track of of semantic features connected to expressions in the lexicon [Pustejovsky, 1991] and allow these feature matrices to be updated as a result of semantic negotiation and coordination subdialogues. Work in this direction may benefit from ideas put forward by [Gärdenfors, 2000], as well as in work on machine learning [Mitchell, 1997] and Latent Semantic Analysis [Landauer and Dumais, 1997]. One version of formal semantics which seems promising for the illumination of concept-level dynamics is the record-type theoretic approach which Cooper has been developing [Cooper, 2005a; Cooper, 2005b]. This formal approach allows for both underspecification or uncertainty of meaning by the use of types of meaning and also a structured approach to meaning analysis which allows for modification of meaning in a way which is not possible, for example, in the classical formal semantics analysis of meaning as functions from contexts to intensions.

## 11 Semantic dynamics and dialogue systems

As we have seen above, vocabularies (and associated meanings) in human-human dialogue are dynamic and changeable. By contrast, a prominent problem in current dialogue systems is that the user must adapt to a static pre-programmed vocabulary. This is related to what Brennan (2000) calls the “vocabulary problem” in spoken dialogue systems. A solution to this problem, quite compatible with what has been proposed here, is sketched by Brennan:

A system could begin a dialogue with a large (an inefficient) vocabulary, in order to allow a user to propose terms. Negotiation - either implicit or explicit - of mappings [...] would be necessary early in the dialogue[...] ... the system would maintain a discourse model of currently active conceptualizations, terms, and mappings to rapidly narrow down the vocabulary its speech recognizer expected. [Brennan, 2000]

Allowing dialogue systems (and other language technologies) to coordinate with users on language use would clearly be beneficial. Relatively simple instances of this would include switching between predefined vocabularies, as well as adding new words to the interpretation lexicon and assigning them a predefined meaning. To do this in interaction with the user would require the system to be able to follow common protocols for coordination of language use. A more complex version is to keep track of of semantic features connected to words in the lexicon [Pustejovsky, 1991] and allow these feature matrices to be updated as a result of semantic negotiation and coordination subdialogues.

Here’s an example of the kind of dialogue that can be envisaged between a user and a city navigation system.

(13) U: What’s the nearest cafe?

S: Danny's deli is just north of your current location. [*using compass + distance register*]

U: uh. I don't know where that is!

S: Turn to face the church. Walk straight ahead for one block, then make a right turn and walk 50 meters. [*using blocks, buildings and streets register*]

U: Oh it's on High Street? [*user switches to street-name ontology*]

S: No. It's on South Bridge, number 45 [*using street-name ontology*]

U: Sorry, I don't know where South Bridge is.

S: Danny's deli is on Clerk Street, near the bridge in the city centre [*using slightly different street-name ontology; South Bridge and Clerk Street are different sections of the same street, but the whole street is sometimes referred to as Clerk Street*]

U: Sorry, I don't get it

S: Turn to face the church and walk along Meadow street towards the north, then at the junction to Clerk Street make a right turn and walk along Clerk Street for about 50 meters until it changes name into South Bridge, then continue on South bridge until you reach number 45. [*combining several registers in one utterance*]

This dialogue shows many of the characteristics of dialogues from the Maze Game experiments [Garrod and Anderson, 1987; Healey, 1997], as well as the Map Task dialogue in Section 4, including switching between and interleaving various registers in response to communication difficulties.

## 12 Conclusion

To account for the observed dynamics of semantic systems in dialogue, this paper has sketched a formalisation of the notion of meaning potential, in the form of dynamic structured usage patterns which are shared within a linguistic community through a process of semantic coordination in dialogue. This process can be described as updates to structured usage patterns resulting from language use. Some basic mechanisms of coordination - feedback, explicit negotiation, and accommodation - have also been outlined.

This paper presents preliminary work aiming towards a unified theoretical account of semantic coordination. Apart from developing the theory and the formal framework further, an important goal is to extend the coverage of this theory by further empirical studies, and to start implementing strategies for semantic coordination in practical dialogue systems.

## BIBLIOGRAPHY

- [Allwood, 1995] Allwood, Jens 1995. An activity based approach to pragmatics. Technical Report (GPTL) 75, Gothenburg Papers in Theoretical Linguistics, University of Göteborg.
- [Bezuidenhout, 2002] Bezuidenhout, Anne L. 2002. Truth-conditional pragmatics. In Tomberlin, James E., editor 2002, *Philosophical Perspectives 16: Language and Mind, 2002*. Blackwell Publishers, Oxford. 105–134.
- [Brennan and Clark, 1996] Brennan, S. E. and Clark, H. H. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition* 22:482–493.
- [Brennan, 2000] Brennan, S. E. 2000. The vocabulary problem in spoken language systems. In Luperfey, S., editor 2000, *Automated spoken dialog systems*. Cambridge, MA: MIT Press. To appear.
- [Briscoe, 2002] Briscoe, E. J., editor 2002. *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge University Press.
- [Clark and Brennan, 1990] Clark, H. H. and Brennan, S. E. 1990. Grounding in communication. In Resnick, L. B.; Levine, J.; and Behrend, S. D., editors 1990, *Perspectives on Socially Shared Cognition*. APA.
- [Clark and Clark, 1979] Clark, E. V. and Clark, H. H. 1979. When nouns surface as verbs. *Language* 55(4):767–811.
- [Clark and Wilkes-Gibbs, 1986] Clark, H. H. and Wilkes-Gibbs, D. 1986. Referring as a collaborative process. *Cognition* 22:1–39.
- [Clark, 1996] Clark, H. H. 1996. *Using Language*. Cambridge University Press, Cambridge.
- [Clark, 1997] Clark, Eve V. 1997. Conceptual perspective and lexical choice in acquisition. *Cognition* 64(1):1–37.
- [Clark, 2003] Clark, E. V. 2003. *First language acquisition*. Cambridge: Cambridge University Press.
- [Cooper, 2005a] Cooper, Robin 2005a. Austinian truth, attitudes and type theory. *Research on Language and Computation* 3(4):333–362.
- [Cooper, 2005b] Cooper, Robin 2005b. Records and record types in semantic theory. *J. Log. and Comput.* 15(2):99–112.
- [Gärdenfors, 2000] Gärdenfors, Peter 2000. *Conceptual Spaces: The Geometry of Thought*. MIT Press, Cambridge, MA, USA.
- [Garrod and Anderson, 1987] Garrod, Simon C. and Anderson, Anthony 1987. Saying what you mean in dialogue: a study in conceptual and semantic co-ordination. *Cognition* 27:181–218.
- [Groenendijk and Stokhof, 1988] Groenendijk, J. A. G. and Stokhof, M. J. B. 1988. Context and information in dynamic semantics. In *Working models of human perception*. Academic Press.
- [Halliday, 1978] Halliday, M.A.K 1978. *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. Baltimore: University Park Press.
- [Healey, 1997] Healey, P.G.T. 1997. Expertise or expertese?: The emergence of task-oriented sublanguages. In Shafto, M.G. and Langley, P., editors 1997, *Proceedings of the 19th Annual Conference of the Cognitive Science Society*. 301–306.
- [Kaplan, 1979] Kaplan, D. 1979. Dthat. In Cole, P., editor 1979, *Syntax and Semantics v. 9, Pragmatics*. Academic Press, New York. 221–243.
- [Landauer and Dumais, 1997] Landauer, Thomas K and Dumais, Susan T. 1997. A solution to plato’s problem: The latent semantic analysis theory of the acquisition, induction and representation of knowledge. *Psychological Review* 104:211–240.
- [Mitchell, 1997] Mitchell, Tom M. 1997. *Machine Learning*. McGraw-Hill, New York.
- [Pickering and Garrod, 2004] Pickering, Martin J. and Garrod, Simon 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27(02):169–226.
- [Portner and Partee, 2002] Portner, P and Partee, B. H., editors 2002. *Formal Semantics: The Essential Readings*. Malden, Mass.
- [Pustejovsky, 1991] Pustejovsky, J. 1991. The generative lexicon. *Computational Linguistics* 17(4):409–441.
- [Recanati, 2003] Recanati, Francois 2003. *Literal Meaning - The Very Idea*. Cambridge University Press.

- [Steels and Belpaeme, 2005] Steels, Luc and Belpaeme, Tony 2005. Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences* 28(4):469–89. Target Paper, discussion 489-529.
- [Traum and Larsson, 2003] Traum, David and Larsson, Staffan 2003. The information state approach to dialogue management. In Smith, Ronnie and Kuppevelt, Jan, editors 2003, *Current and New Directions in Discourse & Dialogue*. Kluwer Academic Publishers.
- [Wittgenstein, 1953] Wittgenstein, Ludwig 1953. *Philosophical Investigations*. Basil Blackwell Ltd.