Improving Coordination on Novel Meaning through Context and Semantic Structure

Thomas Brochhagen

Institute for Logic, Language and Computation
University of Amsterdam
P.O. Box 94242, 1090GE Amsterdam, The Netherlands
t.s.brochhagen@uva.nl

Abstract

Meaning conveyance is bottlenecked by the linguistic conventions shared among interlocutors. One possibility to convey non-conventionalized meaning is to employ known expressions in such a way that the intended meaning can be abduced from them. This, in turn, can give rise to ambiguity. We investigate this process with a focus on its use for semantic coordination and show it to be conducive to fast agreement on novel meaning under a mutual expectation to exploit semantic structure. We argue this to be a motivation for the cross-linguistic pervasiveness of systematic ambiguity.

1 Introduction

Semantic heterogeneity is an inherent aspect of human communication. Nevertheless, successful communication relies on mutual intelligibility. That is, an expression's meaning has to be assumed to be jointly known, or at least be abducible provided other information. Here, the latter communication strategy is addressed. In particular, we focus on the repurposing of an expression to convey novel meaning, derived from the expression's conventional meaning and the context it appears in.[1] As a consequence, single forms may come to be associated with multiple meanings.

We argue such repurposing motivated ambiguity to be driven by two main forces: the predictive power of semantic structure and potential for confounding. On the one hand, using the same expression to convey similar yet non-identical meanings in different contexts allows for the interpretation

of one in terms of the other, modulo context. On the other hand, if the contexts these meanings appear in are either too similar, or too dissimilar, the intended interpretation may fail, leading to suboptimal communication.

Ambiguity in cooperative communication has been argued to be motivated by effort and cost minimization. Santana (2014) shows that ambiguity is evolutionarily advantageous when disambiguating contexts are available and cost is associated with a larger vocabulary size. In a similar spirit, Piantadosi et al. (2012) argue ambiguity to enable a reuse of forms that are easy to produce and comprehend (for example, shorter, phonotactically unmarked, expressions). Thus, according to this view, ambiguity's advantage mainly lies in effort reduction in production while safeguarding comprehension through contextual information.

More generally, the argument is that if context is (at least partially) shared, informative, and cheap, less information needs to be carried by signals. Following Piantadosi et al. this can readily be illustrated by comparing the amount of information required to disambiguate a meaning $t \in T$ with and without context K using Shannon entropy (Shannon, 1948). If K is informative about T, then $H(T) > H(T|K)$. That is, context can alleviate the need for distinct forms for distinct meanings. However, this ignores the subtler issue of how the information of K relates to that of T. In structured domains not all elements are equal: similarity can introduce noise to meaning discriminability or, conversely, emphasize the contrast between dissimilar meanings. Crucially, there are many alternatives ranging from inefficient to efficient contextual exploitation. In turn, this depends on the meaning-form associations of a language and their relation to the contexts they appear in. Other things being equal, an ambiguous language that colexicalizes contextually distinguishable meanings will be more effi-

---

[1] In the following, context is construed broadly as any source of information beyond an expression's literal meaning. It is understood as a condensed prior of the association strength with which an interpretation comes to mind (Franke, 2009).

cient, compression-wise, than one that colexicalizes contextually indistinguishable ones.

The tacit prediction of past research is that languages maximize the utility of ambiguity when colexicalizing meanings that appear in contexts as distinct as necessary to avoid misunderstanding. Thus, if compression and ease of transmission are ambiguity's main driving force, it is not expected for related meanings to be expressed by a single form, as this could make them more prone to be confused. In the following, we argue that ambiguity also has motivations at the semantics-pragmatics interface, where interlocutors may exploit semantic structure to coordinate on novel meaning.

2 Regularities in semantic structure and their relation to context

Assessing the relation between novel meaning, conventional meaning, and the contexts they appear in, presents many difficulties. We begin by considering already conventionalized ambiguous expressions as a proxy for form coexistence of distinct meanings. We do this to support two claims. First, that (at least) some cases of ambiguity in natural language are motivated by semantic relatedness (Apresjan, 1974; Nunberg, 1979; Pustejovsky, 1995).[2] Second, that context and semantic relatedness interact. An in-depth discussion of either claim is outside the scope of the present contribution. However, albeit often presupposed and of certain intuitive appeal, it should be stressed that neither is innocuous.

Semantic relatedness. First evidence for semantic regularities in ambiguity comes from the wide range of genealogically unrelated languages that colexify the same meaning pairs. For instance, the CLiCS corpus (List et al., 2014) lists 297 English noun pairs whose meaning is expressed by a single form in at least 10 languages from three or more language families. For example, 106 languages from 40 families express 'flesh' and 'meat' by a single form. Such cross-linguistic regularities are not expected should an expression's form be ambiguity's main driving force. On a more general level, a number of systematic meaning alternations, such as producer-product, as in *Rembrandt*, or material-artifact, as in *glass*, have also

been attested across multiple languages (Srinivasan and Rabagliati, 2015), although with notably less cross-linguistic coverage.

Furthermore, a body of experimental evidence suggests that the processing of forms that conflate related meanings is distinct from that of unrelated ones (for an overview see Simpson (1984) and Eddington and Tokowicz (2015)). More specifically, semantic relatedness is generally judged as facilitatory for semantic access in comparison to both monosemous and homonymous expressions.

The experiments of Rodd et al. (2012) on the acquisition of novel meaning through the use of forms already associated with conventional meaning are of particular relevance for the claim that reuse of semantic material is conducive to agreement on non-conventionalized meaning. Their results suggest that non-conventionalized meanings are recalled better if they are related to the conventional meaning of a known expression. Similarly, in lexical decision tasks, subjects exhibited increased performance for novel ambiguous words with related meanings but not for unrelated ones. More generally, Srinivasan and Snedeker (2011) show that four-year olds generalize semantic alternations of ambiguous expressions to novel monosemous forms that lexicalize a meaning participating in such alternations. In other words, human interlocutors appear to expect semantic relations to be exploited and generalize known alternations.

Context, disambiguation, and prediction. Contextual information not only has a facilitatory effect on the interpretation of ambiguous expressions (Frazier and Rayner, 1990; Klepousniotou and Baum, 2007). It can furthermore be employed to predict the number of distinct meanings a form has (Hoffman et al., 2013). In particular, distributional semantic models have been shown to provide well-performing context-dependent vectorial representations for the meanings of ambiguous expressions by clustering an expression's co-occurrence counts. Using such methodology, Reisinger and Mooney (2010) found a negative correlation between the variance of cluster similarities and that of human sense annotations: The more similar co-occurrence clusters of an ambiguous form were, the less human raters agreed on their distinct meanings, suggesting an inverse relationship between distributional similarity and semantic discriminability. Boleda

---

[2]A more differentiated classification of ambiguity is not required for the present purpose. Thus, we purposefully avoid referring to polysemy, metonomy, or metaphor explicitly.

et al. (2012) show how distributional models can be used to predict regular meaning alternations for novel words. Here, the similarity of a form's co-occurrence vector to the centroid of two alternation's representations is used to assess whether the form participates in the alternation. As above, this research provides some support to the idea that natural languages do not solely maximize contextual contrast between meanings but that there are regularities between semantic relations and context, reflected in regular colexification patterns.

3   Improving coordination

Taken together, the preceding survey provides indirect evidence for the claim that semantic relatedness plays a role for (at least some types of) ambiguity, as well as for an interplay between interpretation, context, and meaning-multiplicity. In the following, we show that a joint expectation to exploit semantic relations and context leads to improved coordination on novel meaning.

We assume the information provided by context to be shared and noiseless, i.e. interlocutors have access to the same contextual information.[3] Furthermore, we restrict our analysis to cooperative communication. As a consequence, context is taken to be informative about a speaker's intended meaning. The set of meanings compatible with a context k_i, the support of the meaning distribution conditioned on k_i, is denoted by $K_i^\star$, $K_i^\star := \{t | p(t|k_i) > 0\}$. As we are interested in novel use of conventionalized expressions, a fixed message inventory M is considered, where $p(t|m, k_i) = 1$ for exactly one $m \in M$ provided that $t \in K_i^\star$. That is, the messages in M are already conventionally associated with some meanings, guaranteeing communicative success for those meanings. $I(m)$ is the conventional interpretation of a message, $I(m) := \arg\max_t p(t|m, K)$.[4]

So far, when communicating about conventional meaning, interlocutors need not make use of contextual information. Things are different, however, when conveying novel meaning. In such cases, the best a receiver can do is to guess intended t based on the contextual information provided; $p(t|m, k_i) \propto p(t|k_i)$ if $I(m) \notin K_i^\star$. That is, if a message's conventional meaning is ruled out, the best a literal receiver can do is to interpret based on the contextually conditioned meaning distribution. We refer to this communicative strategy as S_l.

Languages that enable strategies akin to S_l are at the stage at which Santana (2014) and Piantadosi et al. (2012) predict ambiguity to be advantageous: whenever T can be partitioned to allow a message to be associated with two contextually disjoint meanings. However, this sidesteps the ad hoc interpretation of such 'surprise' messages in a conventionally incongruent context, as well as the regularities surveyed above. Particularly, it's unclear how meaning can come to be associated with disjoint contexts and whether there are ways to improve this process beyond best guesses.

Under the assumption that there are regularities interlocutors may exploit them to coordinate. The conventional meaning associated with a message can be repurposed in such a way that, in unison with context, a receiver can abduce the intended non-conventional meaning. In accord with the preceding discussion, we assume two factors to play a key role in this process: the relation between the conventionalized and non-conventionalized meanings, as well the information context provides about them. The former indicates the ease to predict or derive one meaning from the other. The latter is a factor for potential equivocation. We call this strategy S_m.

The above can be summarized as follows: Given a context k_i, a meaning to convey t, and a message m, if $I(m) \in K_i^\star$ and $I(m) = t$, then

$$p(t|m, k_i) = 1 \tag{1}$$

If $I(m) \notin K_i^\star$, then

$$p(t|m, k_i) \propto w_1 p(t|R(I(m), t)) + w_2 p(t|k_i) \tag{2}$$

where $R(x, y)$ stands for a relation between x and y, and w_1 and w_2 are weights, $w_1 + w_2 = 1$. The weights control how much import relations have for the non-conventionalized interpretation of a message based on its conventional meaning. S_l corresponds to $w_1 = 0$ and S_m to $w_1 > 0$. Crucially, for a message m, and all meanings t and t' compatible with context k, if $p(t|R(I(m), t)) \geq p(t'|R(I(m), t')) > p(t|k)$, then coordination on t improves for any value of w_1 greater than zero.

---

[3]This assumption is made mainly for expository convenience. As shown by Juba et al. (2011), ambiguity also provides an efficient solution for uncertainty about the degree to which the contextual prior of interlocutors matches.

[4]The conventional interpretation of a message is, generally speaking, independent of a particular context k_i as long as $I(m) \in K_i^\star$.

Thus, S_m can aid coordination on non-conventionalized meaning if (i) there is a relation that appropriately captures the structure of T, and (ii) interlocutors have a mutual expectation to exploit this relation in both production and comprehension. Put differently, S_m has an advantage over S_l in cases where the relation is more informative about the intended meaning than the meaning distribution conditioned on the context. In all other cases performance depends on the value of the weights and the information provided by context.

3.1 Coordination without prior expectation of a particular relation

Prima facie, the above hinges not only on a mutual expectation to use semantic relations to guide coordination, but on the mutual expectation to exploit a particular relation. To see whether coordination improves without this assumption we compare the performance of S_l and S_m in adaptive two-player Lewisian signaling games.

A Lewisian signaling game (Lewis, 1969), $\langle T, M, A, p^*, u_S, u_R \rangle$, consists of a set of meanings T, signals M, and acts A. p^* is a probability distribution over T, and u_S and u_R are the sender's and receiver's respective utility functions. In cooperative signaling sender and receiver have a joint payoff. Thus, a single utility function u can be considered, $u : T \times M \times A \to \mathbb{R}$. Meanings are assumed to be equiprobable, $p^*(t) = \frac{1}{|T|}$, and for each t_i there is exactly one a_j such that $u(t_i, m, a_j) = 1$ if $i = j$. Otherwise, the players receive no payoff. Note as well that a receiver's correct interpretation of a sender's intended meaning is the sole factor influencing the game's outcome. In this sense, meaning-signal associations are arbitrary.

A game iteration begins with a stochastically determined meaning for the sender to convey. To this end, the sender sends a signal. Upon reception of the signal, the receiver selects an act, which in turn determines the players' payoff. Before interacting, sender and receiver have no, or only a partial set of conventions to draw from. Thus, the players' task is to establish a meaning-signal mapping that maximizes their expected utility, i.e. to establish an efficient communication system. To this end, we adopt a common choice for learning in signaling games; Roth-Erev reinforcement learning (RL) (Roth and Erev, 1995). RL pro-

vides a good fit to the behavior of human subjects in comparable tasks (Roth and Erev, 1995; Erev and Roth, 1998; Bruner et al., 2014), is a well-understood learning mechanism, and has convenient convergence properties (Beggs, 2005; Catteeuw and Manderick, 2014). Furthermore, given its simplicity, RL presupposes little sophistication from players.

As with other reinforcement learning algorithms, successful actions in a state of affairs increase a player's propensity for the same action given the same state. More specifically, a player's actions are informed by her accumulated rewards. These are values associated with state-action pairs and represent the success of an action in a given state. In signaling games, states are meanings for the sender and signals for the receiver, and their respective actions are signals and acts. Given a state p, a player will select an action q with a probability proportional to its accumulated rewards, $p(q|p) = \frac{ar(p,q)}{\sum_{q \in Q} ar(p,q)}$. After a game iteration, the accumulated rewards of selected state-action pairs are updated by the players' payoff. As a consequence, a successful meaning-signal-act triple $\langle t_i, m_j, a_k \rangle$ makes a sender more propense to send m_j given t_i in future interactions. Analogously, the receiver is more propense to select a_k given m_j. In this way, players (ideally) learn to communicate efficiently through iterated interactions.

We expand this setup by adding structure to the set of meanings, a set of contexts, as well as two types of players corresponding to S_l and S_m. To add structure, T is modeled as a n-dimensional grid of natural numbers, $T = [o, r]^n$. The relations in T are given by the Manhattan distance between two elements; $R(x, y) := \sum_{i=1}^{n} |x_i - y_i|$. For example, $R((1, 1), (3, 4)) = 5$. These choices were made to accommodate the simple learning and selection mechanisms of the players. In particular S_l receivers proceed by best guesses and only learn through positive feedback. If T were large or continuous it could take a prohibitive amount of time until the first successful action is performed.

The set of contexts K corresponds to all convex subsets of T. That is, if x and y are elements of a context, then either $R(x, y) = 1$ or there is a third element z in the context such that $R(x, y) = 1$ and $R(y, z) = 1$. Consequently, the information about meaning conveyed by a context can be represented by the points it contains. The more elements a context has, the less informative it is. Two extremes

in K are its singletons and the set containing all points in T. The former are contexts where only one meaning is probable and thusly jointly known to be the intended meaning, $p(t|k) = 1$ if $t \in k$ and $|k| = 1$. The latter context is not informative about meaning, $p(t|k) = p(t)$ if $T = k$. More generally, this means that $p(t|k) = \frac{1}{|k|}$ if $t \in k$ and 0 otherwise.

In contrast to classic signaling games, a game iteration now beings with both a meaning to convey, as well as with the determination of a context. While the meaning is a sender's private information, the context is public and shared across all players. In line with the preceding discussion the only restriction we impose is that sampled t has to be an element of sampled k. That is, context never rules out a speaker's intended meaning.

In what follows, we compare the performance of two types of players; S_l and S_m. Both receivers act in accordance to (1) to interpret conventionalized meaning, and (2) for non-conventionalized meaning. They differ in that S_l is given by $w_1 = 0$, whereas S_m corresponds to any value of w_1 greater than zero. The same applies to S_l and S_m senders, mutatis mutandis.

3.2 Simulations

We compare the iterations needed for S_l and S_m players to achieve reasonably efficient communication by means of signals already associated with conventional meaning. Their task is to employ these signals to convey novel meaning. Crucially, players employing S_m begin the game with no bias towards a particular relation to exploit. This means that, while exploration for S_l involves only coordinating on new form-meaning associations, S_m players additionally explore different potential relations.

On the one hand, we expect that once a suitable relation, i.e. one that holds pairwise between all conventionalized and novel meanings, is found, coordination is faster. On the other hand, considering multiple relations, or settling on a relation that does not hold between all pairs, may lead to suboptimal communication and prolong exploration. (Recall that the degree to which relations affect players' choices is controlled by the value of w_1.) Furthermore, it is clear that once a new convention for the (now) ambiguous signals is established, high values of w_1 will interfere with – rather than aid – coordination.
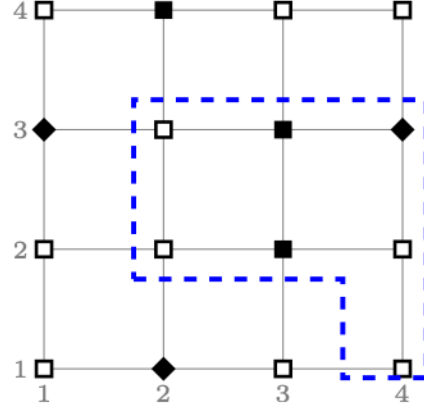

Figure 1: Exemplary instance of an iteration in $[1, 4]^2$. Shapes correspond to meanings. Diamonds are conventionalized and squares are not. Filled out squares are meanings for which players need to establish conventions. The dashed line encloses contextually probable points in this particular iteration.

We compare the effect of different weight values in 100 games of 2000 iterations per value. As mentioned above, $w_1 = 0$ corresponds to S_l. For S_m we consider values for $w_2 \in [.8, .98]$. The set of meanings T is $[1, 4]^2$, yielding 16 potential meanings to choose from, as well as seven distinct relations. Each game is initialized with three randomly sampled meanings taken to be conventionalized and three novel meanings to coordinate on.

The players' performance depends on how many iterations they require to reach an expected utility greater than 0.66 for the latter set of meanings. This corresponds to a better performance than the best suboptimal pooling equilibrium in a signaling game with three meanings, signals, and acts (ignoring the added listener-uncertainty about which three meanings could possibly be intended in the present setup). Reaching this threshold indicates substantial learning as this task is complex for unsophisticated agents. In principle, any element in T could be the intended meaning and learning with RL is slow until at least some successful interactions have transpired. In the worst case, the probability of guessing the right meaning for a receiver using S_l is $\frac{1}{15}$. Figure 1 illustrates an exemplary instance of a single game iteration.

To make the exploitation of relations viable, we ensure that at least one value of the Manhattan distance holds between conventionalized and novel elements. For instance, if points $(1, 3)$, $(2, 1)$ and $(4, 3)$ are conventionalized, and $(3, 3)$, $(3, 2)$ and

78

w_2	mean	SD	Cohen's d
0.95	792	294	3.34
0.90	1238	286	1.34
0.85	1474	289	0.26
0.80	1569	324	-0.08

Table 1: Iterations needed to reach an expected utility greater than 0.66. Cohen's d indicates the difference to the mean of $w_2 = 1$; 1533 (SD = 140).

w_2	mean	SD	Cohen's d
0.95	0.70	0.047	0.49
0.90	0.64	0.054	1.88
0.85	0.60	0.051	2.69
0.80	0.56	0.061	3.19

Table 2: Expected utility after 2000 iterations. Cohen's d indicates the difference to the mean of $w_2 = 1$; 0.73 (SD = 0.04).

$(2, 4)$ are novel meanings to convey, then a distance of 3 allows for their pairwise association. In general, multiple relations hold between conventionalized and novel elements, allowing for more than one relation to be considered. As a consequence, an advantage of S_m over S_l is not certain.

Results & evaluation. In the following, two results are reported. First, the mean of the iterations both types of signalers needed to reach an expected utility greater than 0.66. Second, their mean expected utility after 2000 iterations, indicating long term effects of different w_1-values.

Detailed excerpts of the results, showcasing general trends and the effect size between values of $w_2 = 1$ (S_l) and $w_2 < 1$ (S_m), are shown in Tables 1 and 2, for iterations required and expected utility after 2000 iterations, respectively. Figures 2 and 3 depict plots for all weight values. In the former figure points below the horizontal uninterrupted line show values for which S_m performed better than S_l. In the latter figure points above this line indicate better performance.

Generally, our expectations were met. The higher w_1, the less efficient a communicative system was after a game's conclusion. However, even with respect to expected utility after 2000 iterations, the mean of S_m players was higher than that of S_l players for low w_1-values. For instance, players with $w_1 = 0.02$ reached a mean of 0.76 (SD = 0.023), which is significantly higher than
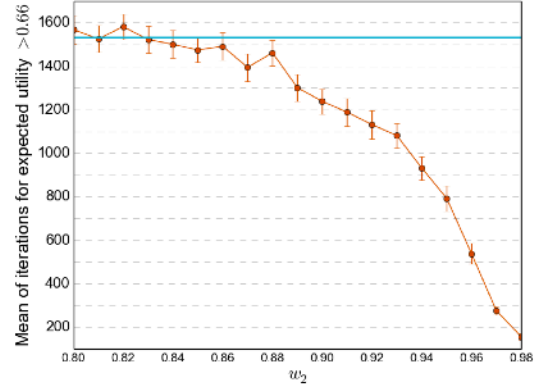

Figure 2: Mean of iterations needed to reach an expected utility greater than 0.66 with 95% confidence intervals. The horizontal uninterrupted line indicates the mean of $w_2 = 1$; 1533 $[1505, 1561]$.
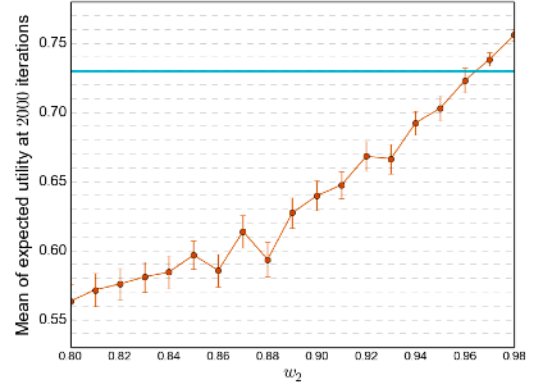

Figure 3: Mean of expected utility after 2000 iterations with 95% confidence intervals. The horizontal uninterrupted line indicates the mean of $w_2 = 1$; 0.73 $[0.717, 0.735]$.

that of $w_1 = 0$ (Cohen's d = -1.11). Crucially, these results show that prior agreement on a single relation is not necessary to uphold the advantage of exploiting semantic relations over best guesses. This is evinced by the range of values that reached the imposed threshold in significantly less iterations than S_l.

In this setup low values of w_1 performed best with respect to learning speed, as well as longer term communicative efficiency. This adds to our previous assumption in that low yet positively valued w_1 improves early exploration without interfering with exploitation. Put differently, a slight bias towards relation exploitation is useful both in short and long term, whereas a major reliance on

this mechanism can have negative effects in the long run, at least when multiple relations are viable candidates.

Overall, even when multiple relations are available, S_m can nevertheless be conducive to fast agreement on novel meaning. This, however, comes at a cost when weights are static. After improving the search for novel meaning, high values of w_1 interfere with further interactions. This is due to the present setup allowing for the "right" relation to hold between more than one of the meanings to convey. As a consequence, S_l generally fared better over time.

4   General discussion

To recapitulate, we argued that repurposing expressions in novel contexts improves coordination when interlocutors exploit semantic regularities. Moreover, our simulations show this advantage to hold without prior agreement on a particular as well. The generality of the latter result, however, is constrained by the setup considered. On the one hand, only a small set of meanings and relations was used. Furthermore, simplifying assumptions were made to model context and its relation to meaning. On the other hand, human agents are able to learn and reason about their interlocutors in more sophisticated ways than our agents, and draw from more information sources. Thus, while its relation to natural language structure and reasoning is tentative, on a more general level the present analysis applies to systems where coder and encoder share an expectation to repurpose information through regular means.

Returning to natural language, our argument partially resembles Grice's modified Occam's razor: "senses should not be multiplied beyond necessity" (Grice, 1978). In a nutshell, Grice argues that, should it be predictable that a speaker would use a particular expression to convey something in a given context, then there is no need to assume this to be a separate meaning of the expression. Without dwelling on the issue whether the meanings considered here constitute novel meanings in their own right – as done so far – the crucial point is that exploiting relations enables predictable interpretation-multiplicity. In this sense, players using S_m can be seen as learning to predict and convey meaning based on the structure of semantic space.

Having a way to predict interpretations, in turn, was shown to lead to faster coordination on non-conventionalized meaning. Furthermore, the longer term comparisons between S_l and S_m suggest that, should the information provided by relations be insufficient to tease apart meaning alternations throughout varying contexts, interlocutors perform best when their choices are only weakly influenced by them. This aligns well with recent research on learning through generalization (O'Connor, forthcoming). O'Connor's results add strength to the claim that generalization speeds up learning whilst paying a cost in precision. Communicatively efficient meaning alternations need to be frequent, and the participating meanings discriminable by the contexts they appear in. In the long run, when potential for confounding exists and high precision is required, interlocutors fare better when coining a new signal for a novel meaning or by drawing from additional information to reduce communicative uncertainty.

We see two main venues for future research. First, there is a need for further analysis involving differently sized and structured meaning spaces, different relations, the addition of noise to the information provided by context, as well as an analysis of population dynamics in larger agent communities.[5] Second, our general proposal requires empirical validation. Here, one possibility is to test its performance on corpus data to predict unwitnessed meaning alternations in a similar spirit to the work of Reisinger and Mooney (2010) and Boleda et al. (2012) surveyed above.

A further issue left undiscussed is that of the cost of ambiguity. In the current proposal cost implicitly came into play as equivocation potential when multiple relations are available for exploitation. Other sources of cost may relate to lexical storage, as assumed by Santana (2014), or processing cost. In particular the latter requires a more detailed treatment. Past experiments suggest ambiguous words with related meanings to be processed faster than monosemous or homonymous words (Rodd et al., 2002; Klepousniotou and Baum, 2007), as well as finer-grained distinctions within their class (Klepousniotou et al., 2008). These aspects relate to issues of lexical storage, lexical representation and lexical access,

---

[5]As noted by an anonymous reviewer, the simplicity of Lewisian signaling games may have to be abandoned to fully explore and expand this proposal. A potentially suitable alternative is given by the language game paradigm as laid out in, for example, Steels (2012).

neither of which were addressed here.

Our overall proposal is based on relations of unspecified nature. To conclude this discussion, we submit that one possibility to model semantic relatedness in a more concrete but framework-independent way is to equate it to transformational complexity between representations, given by the Kolmogorov complexity of one representation conditioned on the other (Chater and Hahn, 1997). Informally, $K(x|y)$ is a complexity measure given by the shortest program that takes y as input and returns x. Kolmogorov complexity is well-understood and widely applicable. Chiefly, it is independent of the representations required for particular applications and provides a good fit for human similarity judgments (see Hahn et al. (2003) for details). Lastly, it addresses the problems of metric-based similarity relations raised by Tversky (1977), who shows that neither triangle inequality nor symmetry need hold for human similarity judgments. The same is true of transformational complexity, as it is compatible with both symmetric and asymmetric relations.

5   Conclusion

Conveying and comprehending novel meaning relies on the interlocutors' mutual reasoning about what is contextually relevant. Among others, meaning can be expressed by composing conventionalized forms, coining new expressions, or by exploiting semantic relations by scaffolding on conventionalized meaning. The present investigation focused on the latter as a communication strategy for fast coordination. We showed that, if a specific relation is mutually expected to be exploited, this mechanism provides a robust solution for reliable and fast coordination. However, when multiple relations are likely candidates, repurposing comes at a risk of lower precision. As a consequence, its advantage depends on the relations available, their regularity across semantic space, previous successful exploitation thereof, and the contexts in which the relevant meanings appear in.

Our analysis draws its main motivation from the cross-linguistic pervasiveness of ambiguous expressions that lexicalize related meanings. In a sense, it is not surprising that certain meaning clusters exhibit systematic alternations. Without risk for confounding, they provide a safe and efficient expansion of a language's expressive range. In other words, relation exploitation provides a partial solution to lexical bottlenecks. Learning and predicting alternations is not only important for our understanding of human communication, but also to overcome analogous bottlenecks faced by computational systems (Navigli, 2009).

More generally, we argued that natural language ambiguity is motivated by more than form-based considerations. When members of a linguistic community are biased towards regularities, repurposing conventionalized material provides an efficient means to convey novel meaning.

Acknowledgments

References

Juri Derenick Apresjan. 1974. Regular polysemy. *Linguistics*, 12(142).

Alan W. Beggs. 2005. On the convergence of reinforcement learning. *Journal of Economic Theory*, 122(1):1–36.

Gemma Boleda, Sebastian Padó, and Jason Utt. 2012. Regular polysemy: A distributional model. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics*, pages 151–160. Association for Computational Linguistics.

Justin Bruner, Cailin O'Connor, Hannah Rubin, and Simon M. Huttegger. 2014. David Lewis in the lab: experimental results on the emergence of meaning. *Synthese*.

David Catteeuw and Bernard Manderick. 2014. The limits and robustness of reinforcement learning in Lewis signaling games. *Connection Science*, 26(2):161–177.

Nick Chater and Ulrike Hahn. 1997. Representational distortion, similarity and the universal law of generalization. In *SimCat97: Proceedings of the Interdisciplinary Workshop on Similarity and Categorization*.

Chelsea M. Eddington and Natasha Tokowicz. 2015. How meaning similarity influences ambiguous word processing: the current state of the literature. *Psychon Bull Rev*, 22(1):13–37.

Ido Erev and Alvin E. Roth. 1998. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American economic review*, 88(4):848–881.

Michael Franke. 2009. *Signal to Act: Game Theoretic Pragmatics*. Ph.D. thesis, University of Amsterdam.

Lyn Frazier and Keith Rayner. 1990. Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language*, 29(2):181–200.

Herbert P. Grice. 1978. Some further notes on logic and conversation. *Syntax and Semantics*, 9:113–128.

Ulrike Hahn, Nick Chater, and Lucy B. Richardson. 2003. Similarity as transformation. *Cognition*, 87(1):1–32.

Paul Hoffman, Matthew A. Lambon Ralph, and Timothy T. Rogers. 2013. Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, 45(3):718–730.

Brendan Juba, Adam Tauman Kalai, Sanjeev Khanna, and Madhu Sudan. 2011. Compression without a common prior: An information-theoretic justification for ambiguity in language. In *Proceedings of the 2nd Symposium on innovations in computer science*.

Ekaterini Klepousniotou and Shari R. Baum. 2007. Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics*, 20(1):1–24.

Ekaterini Klepousniotou, Debra Titone, and Carolina Romero. 2008. Making sense of word senses: The comprehension of polysemy depends on sense overlap. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6):1534–1543.

David Lewis. 1969. *Convention: A Philosophical Study*. Harvard University Press, Cambridge.

Johann-Mattis List, Thomas Mayer, Anselm Terhalle, and Matthias Urban. 2014. CLiCs – database of cross-linguistic colexifications. Version 1.0.

Roberto Navigli. 2009. Word sense disambiguation. *CSUR*, 41(2):1–69.

Geoffrey Nunberg. 1979. The non-uniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy*, 3(2):143–184.

Cailin O'Connor. forthcoming. Evolving to generalize: Trading precision for speed. *British Journal for the Philosophy of Science*.

Steven T. Piantadosi, Harry Tily, and Edward Gibson. 2012. The communicative function of ambiguity in language. *Cognition*, 122(3):280–291.

James Pustejovsky. 1995. *The Generative Lexicon*. MIT Press, Cambridge, MA.

Joseph Reisinger and Raymond J. Mooney. 2010. Multi-prototype vector-space models of word meaning. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 109–117. Association for Computational Linguistics.

Jennifer Rodd, Gareth Gaskell, and William Marslen-Wilson. 2002. Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, 46(2):245–266.

Jennifer M. Rodd, Richard Berriman, Matt Landau, Theresa Lee, Carol Ho, M. Gareth Gaskell, and Matthew H. Davis. 2012. Learning new meanings for old words: effects of semantic relatedness. *Memory & Cognition*, 40(7):1095–1108.

Alvin E. Roth and Ido Erev. 1995. Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, 8(1):164–212.

Carlos Santana. 2014. Ambiguity in cooperative signaling. *Philosophy of Science*, 81(3):398–422.

Claude E. Shannon. 1948. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423.

Greg B. Simpson. 1984. Lexical ambiguity and its role in models of word recognition. *Psychological Bulletin*, 96(2):316–340.

Mahesh Srinivasan and Hugh Rabagliati. 2015. How concepts and conventions structure the lexicon: Cross-linguistic evidence from polysemy. *Lingua*, 157:124–152.

Mahesh Srinivasan and Jesse Snedeker. 2011. Judging a book by its cover and its contents: The representation of polysemous and homophonous meanings in four-year-old children. *Cognitive Psychology*, 62(4):245–272.

Luc Steels. 2012. *Experiments in Cultural Language Evolution*. John Benjamins.

Amos Tversky. 1977. Features of similarity. *Psychological Review*, 84(4):327–352.