Gerhard Schaden
Rapport de mission
Participation au colloque
Formal and Experimental Approaches to Pragmatics

Ce colloque a été organisé par Noah D. Goodman (Stanford), Michael Franke (Amsterdam), et Judith Degen (Stanford), et formait une partie de l'école d'été ESSLLI, qui se tenait à l'Université de Tübingen. Il a réuni une cinquantaine de participants du 11-15 août 2014.

Argumentaire général du colloque (en anglais)

Formal pragmatics has recently experienced a new period of maturation, facilitated by two important factors:

1. the novel application of mathematical modeling techniques, and

2. access to rich experimental data.

The recently emerging field of Experimental Pragmatics has painted a complex picture of the interaction between semantic and pragmatic information in phenomena as diverse as implicature, referentiality, figurative meaning, prosody, and presupposition. In parallel, advances in probabilistic and game-theoretic models that treat pragmatic inference as a problem of reasoning under uncertainty have yielded testable quantitative predictions about the outcome of many different kinds of pragmatic inference. Despite this progress, a great deal of work is needed on the mathematical foundations and quantitative empirical grounding of pragmatics, and, most critically, the connection between the two. The aim of this workshop is to promote dialog and community for these lines of research: strengthening the search for an empirically grounded formal pragmatics.

Given this goal, the workshop aims to provide a forum for advanced PhD students and researchers to present and discuss their work with colleagues and researchers who work in the broad subject areas represented at ESSLLI.

Pour ma contribution aux actes, cf. les pages suivantes.

Markedness, Frequency, and Lexical Change in Unstable Environments

Gerhard Schaden (gerhard.schaden@univ-lille3.fr)
Université Lille 3 & CNRS UMR 8163 STL
Domaine Universitaire du Pont de Bois, 59653 Villeneuve d'Ascq

Abstract

This paper proposes an account of homonymy reversals, based on Polya-urn processes. It assumes that words need to be conceptually grounded, and that the presence or absence of a class of referents may destabilize established meaning differences, as well as cause new meanings to emerge.

The emergence of new meanings is conditioned on an asymmetry of frequency between two words, and requires agents with a Theory of Mind. It will be also argued that markedness cannot be explained by frequency alone.

Keywords: Polya-Urn Processes; Reinforcement Learning; Hyponymy-Reversals; Theory of Mind; Markedness.

Introduction

The lexicon of a language is a relatively unstable domain. Often, words shift their meanings along patterns of polysemy, and what once was a derived meaning becomes the basic meaning of a word. For instance, Modern German *Bein* means 'leg', but in earlier periods, its meaning was just like its English cognate 'bone'.[1] A particularly puzzling phenomenon in this domain is reversals of hyponymy, as exemplified by English *hound* vs. *dog*. In Contemporary English, *dog* is the hyperonym, referring to the totality of *canis lupus familiaris*, but in Middle English and earlier, *dogge* used to be a hyponym of *hound*, which was then the term denoting *canis lupus familiaris*.[2] *Hound* is the cognate of German *Hund*, which is also the form that appears in all other major Germanic languages,[3] and directly derives from Proto-Indo-European *Ḱwon-/*Ḱun-, meaning 'dog' (see Meier-Brügger, Fritz, Krahe, & Mayrhofer, 2002, p. 131). Therefore, English *hound* and *dog* have undergone a reversal of hyponymy between the Middle English and the Early Modern English periods.

Crucially, *dog* and *hound* did not just swap their respective meanings; a *dog* used to be a particular, sturdy kind of *canis lupus familiaris* (see Gąsiorowski (2006, 278) for Medieval textual evidence on this), whereas a *hound* is today a type of *canis lupus familiaris* with good sense of smell destined to be used for hunting.[4]

The question is whether such reversals require first a change in lexical meaning such that they become synonyms, and subsequently rediffentiate, or whether the underlying process is different. I will explore the latter option, and I will argue that pragmatics (i.e., ascribing to an agent the intention to communicate a given meaning with a given form) plays at least in some circumstances a crucial role in the differentiation of meaning.

The general aim of this paper is to investigate a subtype of lexical changes in an unstable environment. More precisely, I will investigate the learning of grounded concepts (see, e.g. Steels, 2008) using Polya-urn processes (see, e.g., Skyrms, 2010) in a context where there is nothing to learn.

A second aim is to investigate the relationship between markedness and frequency. Haspelmath (2006, p. 44) writes that "[...] *since frequency of use seems to explain most of the observed phenomena, we do not need a 'markedness' concept to understand them.*" While I am sympathetic to his general line of argumentation, which tries to show that 'markedness' in linguistics is not a unified category, and aims to substitute more tangible notions for markedness, I think that Haspelmath has gone too far. The basic theoretical claim I will defend in this paper is that one cannot get markedness out of pure frequencies, and that an additional ingredient is required:

(1)   Markedness = frequency of use + pragmatics

Meaning Inference and Reinforcement Learning of Grounded Concepts

I assume following Steels (2008) that words — at least 'simple' words that have obvious and tangible referents like *dog* or *hound* — are conceptually grounded, that is, that they are related to their referents in the surrounding environment. In lexical acquisition, a learner's task is to figure out which kind of entity in his environment a given word refers to. (Lexical) meaning is therefore dependant on the (part of the) world speakers live in.[5]

Support of Meaning Differences and Richness of Environments

When semanticists reason about meaning, they do so generally in abstraction from any particular context, and their terminology reflects this bias. In order to refer unambiguously to the relation between grounded concepts and the environments they occur in — without necessarily making a commitment to the general meaning of a grounded concept —, I will help myself by introducing some definitions.

First of all, (2) introduces a basic relation between meaning and context:

---

[1] This meaning is still residually present today in the anatomical names of many bones, e.g., *Schlüsselbein*, 'collarbone', lit. *key-bone*.

[2] For a discussion of the (rather obscure) etymology of *docga*, see Gąsiorowski (2006).

[3] See, e.g., the dictionary of the Grimm Brothers (see Hildebrand & Wunderlich, 1984).

[4] Cf. http://www.merriam-webster.com/dictionary/hound or https://en.wikipedia.org/wiki/Hound.

[5] This of course is hardly news; the whole *"Wörter und Sachen"* movement (see, e.g. Schuchardt, 1912) is based on this idea.

(2)   A denotational difference between two words *a* and *b* is supported in a given situation *s* and at a moment *t* iff their denotations are not identical (i.e., $[\![a]\!]^{s,t} \neq [\![b]\!]^{s,t}$).

If (2) holds, I will also say that the situation *s* supports a denotational difference between *a* and *b*, or that *s* is sufficiently rich to support a denotational difference between *a* and *b*. In a sense, this is just the inference-side of the familiar phenomenon of interpretation with respect to a model.[6]

Furthermore, if a denotational difference between words *a* and *b* is not supported in a given situation *s*, I will say that words *a* and *b* are *denotation-equivalents* with respect to *s*. This notion needs to be distinguished from the concept of denotational synonymy (see, e.g. Cruse, 2000). Denotational synonyms will be denotation-equivalents for any possible context. However, the reverse is not true: denotation-equivalent with respect to a given situation *s* need not be denotational synonyms, because there may be a meaning difference (encoded in the mind of speakers) that is supported only in a larger context or situation. Thus, the fact that two words are denotation-equivalent in a given situation does not commit us to a position in which they have the same meaning.

Let us look at an example: my Parisian flat does not support a denotational difference between *cat* and *pet*, since the only pet present there happens to be a cat.[7] So, *pet* and *cat* are denotation-equivalents with respect to my flat. However, since all members of the household can access the surrounding environment, where a denotational difference between *cat* and *pet* is supported, my flat is clearly not a linguistically pertinent environment with respect to a possible change in meaning of *cat* and *pet*. Yet, in the long run, if in a given region and for several generations a denotational distinction between words remains unsupported, this could cause a confusion or a shift in meaning between the initially distinct words. Therefore, denotation-equivalents might transform into denotational synonyms, since (human) language learners are situated agents, learning conceptually grounded concepts in the environment they happen to be born into.

At this point, I would like to emphasize that the environments we currently live in are probably not very representative of environments in human history.[8] Currently, the semiotic landscape of language-acquiring children (at least in urban areas in the first world) is potentially rich and, importantly, locally determined to a rather small degree. While the average Central or West European today knows zebras, giraffes, penguins and elephants (none of which live natively in the local environment), he would probably have trouble naming a substantive proportion of the animals and plants living in a three-kilometer radius of his home. In traditional soci-

eties, people are much more dependent in their survival on a good knowledge of local fauna and flora than we are (cf., e.g. Diamond, 2013).

Therefore, modern societies and the natural languages spoken in them are probably less affected by changes in their local, natural environment than (subgroups of) premodern societies. Generally, the more closed and small-scale a linguistic community is, the more likely it is that chance fluctuations in the environment will have an impact on their language's lexicon.

Reinforcement Learning with Polya-Urn Processes

The tool chosen to investigate lexical meaning change is Polya-urn processes, which will be used to represent learning processes relating words to their meanings.

The basic principle of a Polya-urn is that a ball is drawn at random from the urn, and — if the ball corresponds to an appropriate answer — not only will the ball be put back into the urn, but a given quantity of the same type of ball is added to the urn (the *reward*). In this way, the probability of providing an appropriate anwer in the next turn will increase. From a behaviorist point of view, one could say that the appropriate answer has been reinforced. Such a reinforcement process is illustrated in (3):

(3)   a. | URN$_t$ |
             |---------|
             | white:1 |
             | red:1 |

       b. | URN$_{t+1}$ |
             |-------------|
             | white:2 |
             | red:1 |

Let us assume that in (3a), the white ball has been drawn from the urn, and that this is the correct answer. As a consequence, *white* has been reinforced in (3b) by adding an additional white ball to the urn. The addition of a ball increases the probability of drawing a white ball from 0.5 in (3a) to $0.\dot{6}$ in (3b).

If there is only one correct answer, and if no errors occur in reinforcing, the probability of drawing the correct response from the urn at random will approach 1 in the limit. The rate of the increase in the probability of the correct answer depends on the initial inclination weights (that is, the quantity of balls present in the urn in the beginning) and the weight of the reward. The higher the reward with respect to the initial endowment of the urn, the faster the increase in the probability of giving a correct answer. This dependence is illustrated in figure 1.

Simulating Meaning Change With Polya-Urn Processes

For the simulation, I assume two competing words which are initially denotation-equivalents. Therefore, in a signalling game, both words are systematically appropriate answers, and there is in principle nothing to learn. However, the reinforcement process will continue, and modify the weights.

Additionally, I assume that each word has the same number of weighted submeanings — for instance, something like a *qualia-structure* in the Generative Lexicon, as argued for by

---

[6]The idea of support for a given meaning difference could also be extended to diaphasic or diatopic variation, for instance.

[7]In fact, $[\![cat]\!]^{\text{flat(gs)},11/05/2014} = [\![pet]\!]^{\text{flat(gs)},11/05/2014} = $ *Akané*.

[8]Kusters (2003, 2008) makes the same point. He showed that differences in community-structure can even affect morphology, and that what one usually thinks of as external factors may have impact the core of grammar.
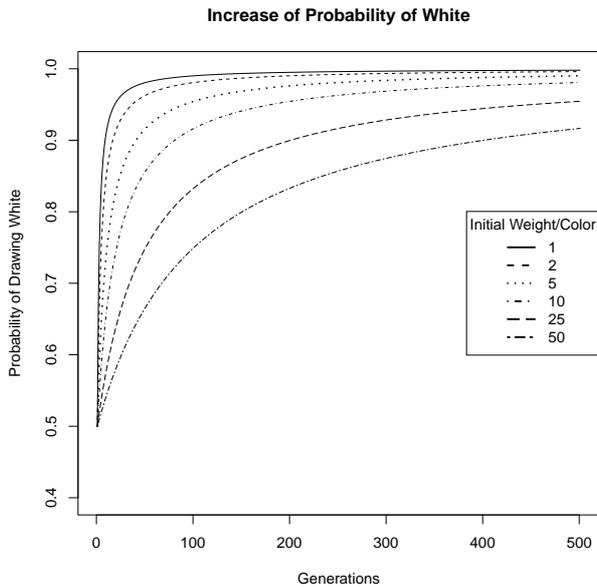
Increase of Probability of White

Figure 1: The increase of the probability of the correct answer in a Polya-urn is initially fast, and then slows down, depending on the initial inclination weight.

Pustejovsky (1995). Remember from the discussion in the introduction that a *dogge* used to be an especially sturdy kind of *canis lupus familiaris*, and that *hounds* today are breeds of *canis lupus familiaris* whose basic purpose is hunting. It happens that in Pustejovsky's qualia-structure, the difference between Middle English *hound* and *dog* can be understood as a specification of *dog* in the *formal*-quale (i.e., shape) which *hound* lacked; and the difference between Modern English *dog* and *hound* can be seen as a specification of *hound* in the *telic*-quale (i.e., purpose or function) which is absent for *dog*.

Dispensing with a monolithic representation for a word may seem to be a costly and unnecessary complication, but it will turn out to be extremely convenient for the simulation, and there are other factors that favor it. First of all, the fact of having independent submeanings allows us to bridge the difference between lexical and encyclopedic knowledge, which may not be all that useful when working with grounded concepts. Second, linguistically, having meanings that are more structured facilitates dealing with certain issues of compositionality.[9] Finally, there is some evidence from functional dissociation in brain-damaged patients that the parameters of shape and function (these are the crucial ingredients distinguishing *dog* from *hound*) are treated by different regions of the brain (as reported by Bermúdez, 2005, p. 20).

Technically, each word is modeled as a structure (or an object) where the initial inclination-weights n are the same for each submeaning, and where each submeaning has exactly

one type (for instance, regarding the *telic*-qualia of the generative lexicon, there is exactly one function for the object). All weight is awarded to some submeaning; there is no global weight. At the beginning, we thus have the situation sketched in table 1:

Table 1: Structure of Words in the Simulation.

	Word1	Word2
submeaning1	type-1: n	type-1: n
submeaning2	type-1: n	type-1: n
submeaning3	type-1: n	type-1: n
submeaning4	type-1: n	type-1: n

Nature chooses at random one submeaning, and the speaker[10] chooses one word according to its relative weight for that submeaning. For instance, if Word1 has 10 balls at submeaning1, and Word2 5 balls, and if Nature chooses submeaning1, then the speaker will choose Word1 with probability $\frac{2}{3}$, and Word2 with probability $\frac{1}{3}$. Since according to our assumption the words are denotation-equivalents, whatever word is drawn will be reinforced according to some preestablished reward. Therefore, we have 4 independent Polya-urns (one for each submeaning), each containing balls according to the weights of each word at the corresponding submeaning.

Figure 2 illustrates a sample outcome of the reinforcement of the two words. It represents the outcome of 4000 rounds of reinforcement, where the initial inclination weight was 1, and the reinforcement reward also corresponded to 1. One can see that in this particular case, for each submeaning (referred to as "Quale" in figure 2), one word has been reinforced more often than the other.

As illustrated by the leftmost boxplot in figure 3, the outcome in figure 2 is rather typical for this kind of initial inclination weight and reward: the median difference corresponds roughly to 500.[11] This means that, for approximately 1000 iterations per submeaning, we will obtain as a median a weight of 750 for one word, and 250 for the other.[12] However, there are also cases where the difference is close to 0, that is, where both words have nearly the same weight for a submeaning, and cases where the difference is close to 1000, that is, where one word has (nearly) always been reinforced for a given submeaning.

Figure 3 as a whole shows however that this result depends crucially on the relationship between initial inclination weight and the reinforcement reward: the higher the incli-

---

[9]See, for instance, the discussion of default readings of verbs by Pustejovsky (1995, p. 88f.).

[10]I will talk about the *speaker*, but it should be noted that this is simply a signalling agent, which is not necessarily human. In any case, the signaller chooses the signal according to the weights in the urn, without any other consideration.

[11]In fact, the precise median here is 495.5.

[12]I have plotted in figure 3 the absolute difference in weight between the two words with respect to a given submeaning k (i.e., **abs**(weight of submeaning k of Word1 $-$ weight of submeaning k of Word2)).

Sample Outcome of Reinforcement of Submeanings

Absolute Differences in Submeanings

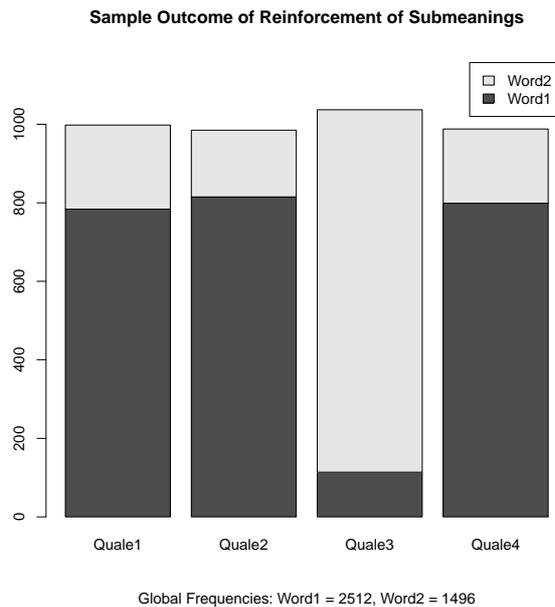Global Frequencies: Word1 = 2512, Word2 = 1496

Figure 2: Sample result of reinforcement learning of denotation-equivalents, after 4000 iterations, with initial inclination weights of 1 and a reward of 1.

Figure 3: Absolute Differences Between Corresponding Submeanings With Differing Initial Inclination Weights

nation weight, the lower the absolute difference between the corresponding submeanings. Whereas the median difference between the weights of corresponding submeanings with an inclination weight of 1 amounts to 495.5, the same median is down at 98 with initial inclination weight of 25, and at 72 with an initial inclination weight of 50.

Let me summarize our findings so far: given low initial inclination weights, the assumed circumstances — that is, reinforcement learning in contexts where there is nothing to learn — will produce as a rule strongly unbalanced weights in a given submeaning. Therefore, we have shown that at least under some circumstances, a purely stochastic process is able to provide differing frequencies of use for two words, in something that one can see as a very primitive kind of text.

Coming back to the quote from Haspelmath (2006), equating (textual) frequency with markedness, does this mean that we have derived something like markedness in our simulation? For instance, in our sample outcome depicted in figure 2, could we say that Word2 has become the marked alternative, whereas Word1 is the unmarked one? Or would we have to be more prudent, and say that Word1 is the unmarked alternative for submeanings 1, 2 and 4, whereas Word2 is the unmarked variety for submeaning 3?

The answers to these questions depend on what exactly one understands by 'markedness'. I would argue that in our case, we do not have semantic markedness in any meaningful sense. While markedness is a tricky (and often unclear) notion (cf. again Haspelmath, 2006), it normally involves the *exploitation* of frequency in order to convey differences or
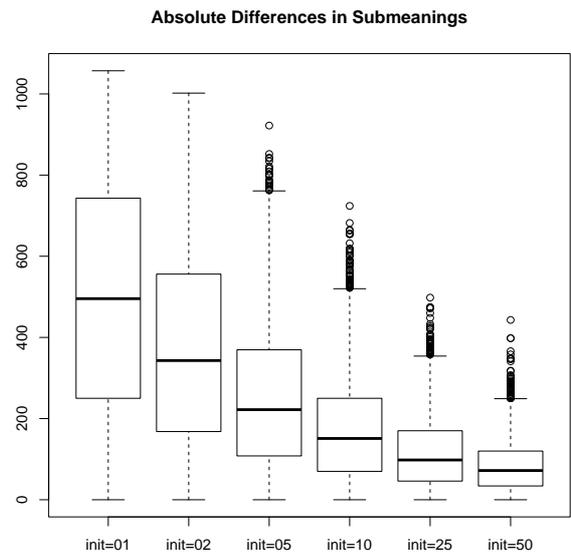
specializations in meaning. What we arrived at in the simulation is simply a difference in frequency with respect to given submeanings. For instance, for signalling submeaning Quale1, an agent would choose Word1 with a probability of roughly 0.8, and Word2 with a probability of roughly 0.2. But in the present setup — which is purely stochastic —, both words continue to have the same lexical meaning, and I do not think that there is any basis for claiming that either word has acquired any specialization.

That being said, let us now have a look at factors that might cause a differentiation in meaning.

A Change in the Environment

What happens if the world changes in a way that might affect the denotation of the word-pairs? Sticking with Pustejovsky's qualia-model, let us assume for instance that a new function for the denoted object arises, but that the other submeanings are not affected by this change. For instance, assume that through the introduction of a new breed and as a consequence of a change in fashion, dogs are used not merely for hunting as before, but also as lap dogs. Yet, assume that the newly introduced entities clearly qualify as *canis lupus familiaris*, and therefore, fall under the denotation of both *hound* and *dog*.

Technically, the process described above amounts to a mutation in the environment, which causes also a mutation in one submeaning. As a consequence, we will need to develop two subtypes in some quale, as in (4) (where k and i denote weights associated to the different types of the submeanings).

(4)   submeaning n: $\begin{cases} \text{type-1:} & k \\ \text{type-2:} & i \end{cases}$

Will this have an impact on the two words that were denotational synonyms before? I will show that this depends on at least two factors: first of all, the strategy chosen for relating (or not) the new submeaning to the old submeaning, and second, the respective weights of the two words before the change in the world takes place.

Agents With Or Without Theory of Mind

If the new submeaning is treated as being entirely independent from the old submeaning (and therefore, the new submeaning is set up as a new Polya-urn with no connection whatsoever to the old submeaning), there will be no tendency for one of the two words to specialize for the old, and one of the two words to specialize for the new submeaning — just as there was no tendency to recruit one word for a particular submeaning before. I suggest treating this as the default case, since it does not require anything that would be different from the procedure used up to now, and merely adds one Polya-urn. Therefore, this should be the base case if one assumes non-intelligent (or: 'mind-blind', cf. Baron-Cohen (1997)) agents. Of course, there is nothing that would prevent the new submeaning from acquiring — through stochastic reinforcement — a frequency-pattern that is the inverse of the old submeaning. However, this will not happen systematically, and there is no reason whatsoever why one word should specialize for one submeaning, and the other word for the second.

What would it take then to cause a differentiation in meaning? The crucial dimension is to establish a link between the old type of the submeaning and the new one. Human agents might take into account the fact that before the change of the world, one of the two words was strongly correlated to the (old) subtype of the mutated submeaning, and assume that the hearer will take this into account. This pattern of reasoning is standardly assumed in Bi-directional OT (cf. Blutner, 1999) in order to derive markedness patterns in natural language. The predicted outcome of such a process is that using the word that is less correlated with the old type will be preferably used in order to signal the new meaning — if there is an incentive for speaker and hearer to do so, which may be the case if there are differential payoffs for using one entity or the other for the two functions. For instance, hunting with a chihuahua might reduce the chances of finding game, and a mastiff is likely to make a wearisome lap-dog.

It is important to notice that the differentiation in the meanings of the two words with respect to the newly arisen type of submeaning is pragmatically induced, and involves — at least in the version of Bi-directional OT — reasoning about (or a simulation of) what the other participant in the communication would do. In brief, the differentiation process presupposes agents with a Theory of Mind, who exploit the difference in frequencies between the two linguistic forms in order to convey a particular meaning.

Relating Subtypes of Meaning

Once we have concluded that the two meanings are to be put into relation, the question is how exactly we should do that. Up to now, frequencies and random draws have played an important role, and it is desirable not to lose these properties, and to go deterministic merely due to the presence of a second choice. Furthermore, I assume that if there was a strong separation in the old type of the submeaning, this would involve a strong association of one word with the old meaning — which in turn should give incentives to associate the new meaning with the other word, even if both words have identical initial inclination weights for the new type. However, if the weights at the old submeaning are roughly identical, the basis for deciding which word should be associated with which submeaning will be much less clear. These are the basic desiderata that I have tried to incorporate into the algorithm.

The formula I used for weighing off frequencies in case of a submeaning having two subtypes is thus the following:

(5)   The probability for choosing a word W for a subtype M of a submeaning Σ, given

   a.   a word V being a denotation-equivalent of W;

   b.   a second subtype V of the same submeaning Σ,

   is:

$$\frac{\dfrac{\text{weight of } M \text{ in } W}{\text{weight of } N \text{ in } W}}{\dfrac{\text{weight of } M \text{ in } W}{\text{weight of } N \text{ in } W} + \dfrac{\text{weight of } M \text{ in } V}{\text{weight of } N \text{ in } V}}$$

Assume for the sake of the argument that we want to calculate the weight for the use of Word1 for meaning *subtype-1* of some submeaning k, given the weights in table 2 (assuming thus that meaning *subtype-2* has never been reinforced). Filling the values of 2 into formula (5) gives us $\frac{800/1}{800/1 + 200/1} = 0.8$.

Table 2: Sample Case of Submeaning Mutation after Change of the Environment.

		Word1	**Word2**
Submeaning k	subtype-1	800	200
	subtype-2	1	1

So far, this does not change from what we have experienced before. However, if we calculate the probability of choosing Word2 for the new meaning, the probability is not 0.5 — as would be the case if subtype-2 was independent from subtype-1 —, but rather 0.8 (since we have $\frac{1/200}{1/200 + 1/800} = 0.8$). And if Word2 is chosen and reinforced, this probability rises to $0.\overset{\cdot}{8}$.

Assuming that such a pragmatic reasoning takes place (and it does not matter what exactly the chosen algorithm will be to weigh off the two submeanings), it will only have an effect if the submeaning undergoing change is associated strongly enough with one word. If the frequencies of the two words are

Changed World and Pragmatics: Density Estimates for Different Initials
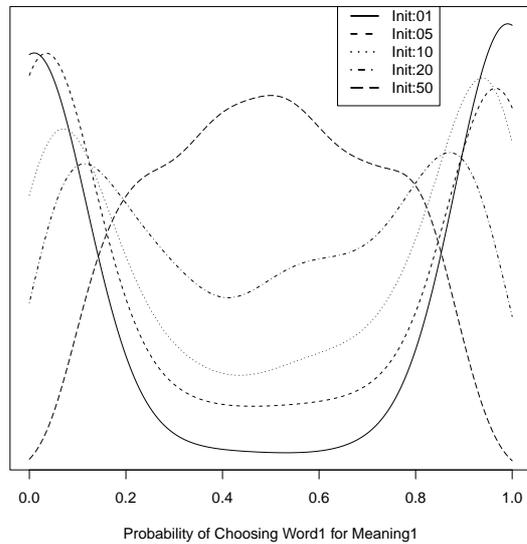

Figure 4: Variation of initial inclination weights and the impact of pragmatics: clusters around 0.5 indicate free variation; clusters around 0 and 1 indicate a separation of the meanings of the two words.

roughly equal before the change of the world, the pragmatic procedure will fail to achieve the separation of the meaning of the two words. This can once again be shown by varying the different initial inclination weights of the Polya-urns: the higher the initial endowment, the less the differentiation into two clearly separated meanings will be noticeable. Figure 4 illustrates this behavior. I have plotted 1000 simulations involving a mutation of the world for varying initial inclination weights, and the diagram shows density estimates of the probability for choosing Word1 for the subtype1 of the mutated submeaning. If the probability mass clusters around 0 and 1 (with hardly any cases in between), this means that either Word1 will be chosen (nearly) all the time or hardly ever for that particular meaning — which is what happens with low inclination weights. Such a pattern is evidence that the words have acquired a specialization in meaning. However, if the probability mass clusters around 0.5 (as can be seen with the inclination weight of 50), this indicates that there is a tendency to obtain free variation: at 0.5, one or the other word is chosen for one or the other meaning with the same frequency.

The reason for this pattern is obvious: the lower the initial inclination weight, the higher the mean difference before the change of the world (see figure 3, page 4). And the higher the difference before the change of the world, the stronger the relative associations with one particular word.
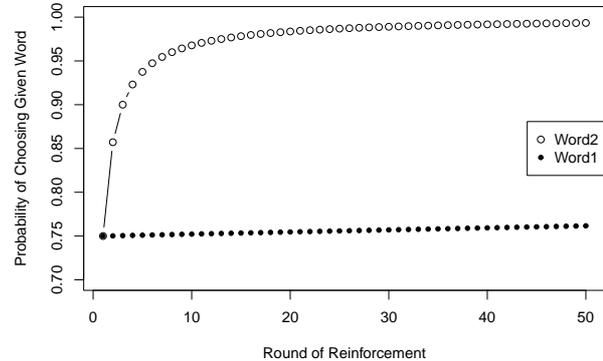
Differing Effects of Reinforcement by 1

Figure 5: Reinforcement by the same quantity gives drastically different results if the new or the old subtype is reinforced. Notice the sharp initial increase, followed by the flattening out of the curve.

Further Consequences of the Chosen Formula

As figure 4 shows, formula (5) has in some circumstances a bias for separating the meanings of the two subtypes in the mutated quale. I will explain the origin and exact working of this behavior now.

Let us assume a basic configuration with a reward of 1, and a configuration just like in 2, but with Word1 having a weight of 750 at subtype-1, and Word2 a weight of 250 at subtype-1 (that is, the median outcome for an inclination weight of 1), as illustrated in table 3. As the reader may check, the probability of choosing Word1 for subtype-1 will be 0.75, and the probability of choosing Word2 for subtype-2 will equally be 0.75.

Table 3: Sample Case of Submeaning with Two Subtypes.

		Word1	**Word2**
Submeaning k	subtype-1	750	250
	subtype-2	1	1

Reinforcing Word1 for subtype-1 and reinforcing Word2 for subtype-2 will not however have the same consequences: If Word1 is systematically reinforced for 50 rounds for subtype-1, but subtype-2 is never drawn by Nature, the probability for drawing Word1 for subtype-1 will have risen from 0.75 to a little more than 0.76; if Word2 is systematically reinforced for 50 rounds for subtype-2, but subtype-1 is never drawn by Nature, then the probability of drawing Word2 for subtype-2 will have risen from 0.75 to a little more than 0.99. The threshold of 0.99 is already reached after 33 rounds. This effect is illustrated in figure 5, which shows once again that the increase in probability for Word2 is very steep at first, and than flattens out gradually.

So, the bigger the difference in frequency of the two words with respect to the submeaning before the mutation, the smaller the probability that the strongly associated word will ever be reinforced for the new subtype of meaning.

Notice however, that according to formula (5), the first reinforcement of the new submeaning will have drastic consequences, whatever its direction: should in our case Word1 be reinforced for subtype-2 (which would happen with a probability of 0.25), the probability of choosing Word1 for subtype-1 will go down to 0.6.

A last important property of the formula is that it has symmetry built in, and a bias for separation. That is, if Word1 is reinforced for subtype-1, this will alter at the same time, and by the same amount, the probability of choosing Word2 for subtype-2, and vice-versa. Assume that in the scenario outlined in table 3, Word2 is reinforced for subtype-2 to 2, with everything else remaining as is. This reinforcement increases of course the probability of choosing Word2 for subtype-2 from 0.75 to 0.857. But at the same time, it also increases the probability of choosing Word1 for subtype-1 from 0.75 to 0.857.

A (But Not *The*) Scenario of Hyponymy Reversal

Now we have all the ingredients in place to sketch a scenario of how the hyponymy reversal between *dog* and *hound* (or any other two words) might have happened.

A first step would have to be an impoverishment of the environment, rendering the two words denotation-equivalents. Next, over the rounds of reinforcement learning, the submeaning whose mutation would eventually be caused by a change in the environment would have to receive a strongly biased frequency in favor of one particular word. Then, after a second change in the environment, the previously less favored word would become associated with (some aspect of) the newly introduced type of referent through rounds of (pragmatically conditioned) reinforcement learning.

Technically, I have not yet derived the development of a relation of hyponymy, since both forms remain in principle possible for all meanings — although some will become less and less likely to appear for a given form. The solution to this problem will probably need to involve *forgetting* one subtype (or one submeaning). Exploring the impact and dynamics of forgetting cannot be done in any detail here for want of space (but see Skyrms, 2010, p. 133ff. for a discussion of forgetting in reinforcement learning). Suffice it to say that the smaller the weight of a subtype, the higher its risk of suffering elimination by forgetting. Intuitively, a subtype with little weight is a subtype that has not been encountered often (if at all) in connection with a given word, and all things being equal, it seems reasonable that forgetting affects the rare rather than the very frequent.

We have seen in the case-study above that the probability of using Word2 for the established subtype is low, and declines rapidly. Therefore, it is a good target for forgetting, consequently rendering it impossible for use for the established subtype. The net result would be that Word2 specializes for one subtype of a quale, whereas Word1 continues to be appropriate with both subtypes (although it would be used more often with the established subtype). In this way, Word2 becomes a hyponym of Word1, specifying information for a submeaning that Word1 does not specify.

Summing up, the scenario involves the impoverishment of the environment, causing words whose denotations overlapped (that is, words that had been partial synonyms according to the definition of Cruse, 2000) to become denotation-equivalents. Then, reinforcement learning ensues, followed by a second change in the environment reintroducing diversity, and finally, forgetting.

Before concluding, it needs to be stressed that randomness plays a major role in the simulation. Even with identical external environments, there is no guarantee whatsoever that meanings will shift (or not) in one way or another. This is probably a welcome fact. While I have not found comparative studies of populations of *canis lupus familiaris* in different European countries, we cannot assume that Medieval England (where a hyponomy reversal took place) was very different from Medieval Germany or Denmark (where nothing of that sort occurred). So, we should look for models that make such a shift possible, but not necessary.

Conclusion, And A Puzzle In Perspective

In this paper, I have simulated the effect of reinforcement learning of conceptually grounded words in an unstable and delimited environment. The simulation used Polya-urn processes on internally differentiated meanings. Since changes in the extra-linguistic environment were crucial for the proposed scenario, the particular brand of hyponymy reversals explored here is therefore rather an instance of external change.

I have also insisted on the necessity of pragmatic processes, and more precisely, the necessity of agents able to guess what another agent might infer from their signal, for the (re-)differentiation of meaning. Against the position of Haspelmath (2006), I have argued that markedness cannot be derived from frequency alone, in the absence of agents disposing of Theory of Mind to interpret these frequencies.

It is obvious that not all changes of lexical meaning can have their source in a changing environment. For instance, concerning the meaning shift of German *Bein* from 'bone' to 'leg', it is nearly unimaginable to have a substantial community of Human agents lacking systematically either legs or bones over a prolonged period. Thus, a necessary future direction of research is to explore under what circumstances a differentiation of meaning is achievable when there is some, though only limited, denotational support to the meaning difference of two words.

At least one important subclass of such cases turn out to be stag hunts (for a book-length presentation, see Skyrms, 2004). This is notably the case when the two words at stake are hyponyms. In a stag hunt, players have the option of going for a zero-risk, low-benefit hare, which they will obtain

irrespective of what the other player does. However, a player might also try to capture the more risky, but more rewarding, stag — which requires however cooperation from the other player. A sample payoff matrix for a game of stag hunt is given in table 4.

Table 4: Sample Stag-Hunt Game.

	Stag	Hare
Stag	2,2	0,1
Hare	1,0	1,1

Assume now a situation with reinforcement learning like we had before, but where one can use either a hyponym (e.g., *cat*) or its hyperonym (e.g., *pet*), and where the denotational difference is supported (that is, where *n%* of pets are cats, and 100-*n%* are pets, but not cats). Rewards are not equal, but follow table 4. In a game where one has to match either the hyponym or the hyperonym, using the hyperonym is the safe play (and corresponds thus to the hare), since there will be no risk with it.[13] The hyponym may yield a greater payoff, but it also involves a greater risk (it corresponds therefore to the stag).

Now comes the puzzling part: it is not clear (at least to the author of these lines) how the structuration of lexicons with hyponymy is sustainable given this fact. As far as I am aware, conditions under which a stag-playing strategy can emerge and persist do not apply in cases of lexical signalling (e.g., locality constraints between players, cf. Skyrms, 2004, p. 15ff.). The fact however that hyponymy is one of the central organizational principles of natural-language lexicons remains empirically obvious.

Acknowledgements

References

Baron-Cohen, S. (1997). *Mindblindness. An essay on autism and theory of mind.* Cambridge, MA: MIT Press.

Bermúdez, J. L. (2005). *Philosophy of psychology. A contemporary introduction.* London: Routledge.

Blutner, R. (1999). Some aspects of optimality in natural language interpretation. In H. de Hoop & H. de Swart (Eds.), *Papers on optimality theoretic semantics* (pp. 1–21). Retrieved from `http://www.blutner.de/optimal.pdf`

Cruse, D. A. (2000). *Meaning in language. An introduction to semantics and pragmatics.* Oxford: Oxford University Press.

Diamond, J. (2013). *The world until yesterday: What can we learn from traditional societies?* London: Penguin.

Gąsiorowski, P. (2006). The etymology of Old English **dogca. Indogermanische Forschungen*, *111*, 276–284. doi: 10.1515/IDGF.2006.015

Haspelmath, M. (2006). Against markedness (and what to replace it with). *Journal of Linguistics*, *42*, 25–70.

Hildebrand, R., & Wunderlich, H. (Eds.). (1984). *Deutsches Wörterbuch von Jacob und Wilhelm Grimm.* München: dtv. Retrieved from `http://urts55.uni-trier.de:8080/Projekte/DWB`

Kusters, W. (2003). *Linguistic complexity. The influence of social change on verbal inflection* (Unpublished doctoral dissertation). Universiteit Leiden, Leiden.

Kusters, W. (2008). Prehistoric and posthistoric language in oblivion. In R. Eckardt, G. Jäger, & T. Veenstra (Eds.), *Variation, selection, development. Probing the evolutionary model of language change* (pp. 199–218). Berlin: Mouton de Gruyter.

Meier-Brügger, M., Fritz, M., Krahe, F., & Mayrhofer, M. (2002). *Indogermanische Sprachwissenschaft* (8th ed.). Berlin: Walter de Gruyter.

Pustejovsky, J. (1995). *The generative lexicon.* Cambridge: MIT Press.

R Core Team. (2014). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org`

Schuchardt, H. (1912). Sachen und Wörter. *Anthropos*, *7*, 827–839. Retrieved from `http://schuchardt.uni-graz.at/werk/schriften/gesamte-edition/online/493`

Skyrms, B. (2004). *The Stag Hunt and the evolution of social structure.* Cambridge: Cambridge University Press.

Skyrms, B. (2010). *Signals. Evolution, learning, & information.* Oxford: Oxford University Press.

Steels, L. (2008). The symbol grounding problem has been solved. So what's next? In M. de Vega, A. Glenberg, & A. Graesser (Eds.), *Symbols and embodiment: Debates on meaning and cognition* (pp. 223–244). Oxford: Oxford University Press.

---

[13]Or, at least, less risk: where the hyperonym is a false answer, the hyponym will by definition be a false answer as well. The reverse is however not true.