MODELLING METAPHOR INTERPRETATION THROUGH INFORMATIVENESS IMPLICATURES AND EXPECTATIONS OF NORMALITY

Bruno L. O. Andreotti
University of Toronto

The goal of this paper is to demonstrate some preliminary work that suggests that metaphorical interpretation can be understood as the product of a general contextualization mechanism that determines which of the various properties associated with a stereotype – its *lexical defaults* – are available to be contributed to the discourse as informativeness implicatures (Levinson 2000) at the moment of utterance. I argue that a simple but non-monotonic contextualization algorithm that suppresses any lexically encoded stereotypal information that is incompatible with or irrelevant to the topic of discussion (allowing the remainder to "pass through") automatically produces metaphorical interpretations if it is allowed to similarly suppress core semantic/literal content.

To this end, in Section 1 I discuss lexical defaults and the approach I employ to formally approximate them based on Yalcin's (2016) expectations of normality. In Section 2 I introduce the contextualization algorithm in question as it applies to generating informativeness implicatures of utterances interpreted literally. In Section 3 I demonstrate that allowing this algorithm to suppress core semantic content leads to it correctly predicting metaphorical discourse contributions.

1. Lexical defaults

My approach to lexically encoded stereotypal information is to approximate it using lists of presumptive properties which I call *lexical defaults*. These are usually properties that the referents of a term generally but not universally possess. For example, by default, if we are told that *Tammy is a cat*, barring accepted information to the contrary in the common ground of the discourse we automatically infer that *Tammy has a tail*. However, Tammy may lack a tail, having either lost it in some unfortunate situation or having been born without one, and still be a cat (albeit a somewhat atypical one). Lacking a default does not preclude membership in a class whose stereotype expects it. I employ three tests to diagnose defaulthood:

- **Conjunction Test**
  The conjunction *but* is licenced with the negation of a default:
  - Tammy is a cat **but** she has no tail.
    ?? Tammy is a cat **but** she has a tail.

- **Weak Necessity Test**
  The modal *should* is acceptable with the affirmation of a default, while *must* is not acceptable:

Actes du congrès annuel de l'Association canadienne de linguistique 2022.
Proceedings of the 2022 annual conference of the Canadian Linguistic Association.

- If Tammy is a cat, she **should** have a tail.
  # If Tammy is a cat, she **must** have a tail.

- **Generic Test**
  Defaults can be quantified generically but not universally:
  - Cats have a tail.
    # All cats have a tail.

Broadly speaking, defaults can be divided into three categories: *inductive*, *pre-emptive*, and *arbitrary* defaults. Inductive defaults arise from properties that most instances of a class possess, such as *cats have a tail* or *birds have feathers*. Pre-emptive defaults are properties which may not apply to the majority of the class, but are still pre-empted of the class in question (often because of a danger it poses), such as *spiders bite* or *plastic bags suffocate animals.* Arbitrary defaults, the most striking category, may have no basis in reality whatsoever, such as *snakes are deceptive* or *owls are wise*. These retain a cognitively dissonant acceptability within the three diagnostics outlined above even after it has been pointed out that they are nonsensical:

(1) a) ! Hissy is a snake **but** she is not deceptive.
b) ! If Hissy is a snake, she **should** be deceptive.
c) ! Snakes are deceptive.

Lexical defaults are drawn from the encyclopedic common-sense knowledge of a speech community. While they cannot be said to be strictly speaking *semantic* – they do not have any obvious effect on the truth conditions of utterances interpreted literally – they are also not entirely *pragmatic* because they are arbitrary, partially context-independent, and associated conventionally to specific words/phrases (or specific senses thereof).

The fact that they are arbitrary can be observed in arbitrary defaults outlined above, as well as the various defaults present in social stereotypes that underlie various forms of systemic prejudice. Their lexical conventionality can be seen from cases where semantic synonyms carry different defaults, e.g. *mob* (angry, dangerous) vs. *crowd* (disorganized, crammed), *strange* (curious, intriguing) vs. *weird* (repulsive, undesirable).

Nevertheless, the standard position to take on this phenomenon is to approach them as generalized informativeness implicatures in the sense of Levinson (2000). Levinson proposed three heuristics that enrich the semantic content of utterances, each capable of generating cancelable implicatures:

- **Quantity Heuristic**
  What is not said is not the case.
- **Informativeness Heuristic**
  What is simply expressed is stereotypically exemplified.
- **Manner Heuristic**
  What is said in an abnormal way is not normal.

The Quantity Heuristic generates well studied scalar implicatures; e.g. *some people own a car* implicates that *not every person owns a car* despite the sentence being semantically true in a situation where everyone does. Similarly, the Manner Heuristic generates well studied markedness implicatures; e.g. *John opened the door* is logically equivalent to *John caused the door to open*, but the atypical prolixity of the latter expression leads to the implicature that the event itself was somehow atypical.

However, the implicatures generated by the Informativeness Heuristic are not as well explored. Some examples of informativeness implicature are not clearly derivable from stereotypes. For instance, (2) being interpretable as (2a) but not (2b) despite being logically compatible with either seems to be a conventional product of event conjunction or a general product of discourse coherence in narrative descriptions of events.

(2) Amy drove to the supermarket **and** visited the pharmacy.
a) Amy drove to the supermarket **and then** visited the pharmacy.
b) ?? Amy visited the pharmacy **and then** drove to the supermarket.

Other examples are more clearly derived from stereotypes, such as *road* implicating *hard surface* (Levinson 2000). Focusing on these examples, it is then not clear that they are calculable as generalized implicatures. We can easily describe a generalized calculation from *(it is a) road* to *(it is a) typical road* as an anti-markedness implicature[1], but we cannot then extend this to *(it has a) hard surface* without reference to a source of conventional stereotypal information that is external to the calculation that specifies what to expect from a typical road.

Lexical defaults thus differ from informativeness implicatures in that they are the residents of this external source of stereotypal information. Informativeness implicatures are *drawn from* lexical defaults, which exist independently.

1.1 Modelling lexical defaults

Stereotypal information is not fully capturable with standard logical formalism – the stereotype of a *rose*, for instance, specifies fine grained information about how a rose should smell that is difficult if not impossible to express precisely even in natural language, despite the fact that we can easily tell when the smell of a particular rose deviates from our expectations. However, the effect of stereotypal information on natural language interpretation is *approximable* through logical formalism.

Based on their connection with modal weak necessity, I formulate lexical defaults as localized expectations of normality (Yalcin 2016; see also Kratzer 1991). Yalcin specifies a partial preorder $\succsim_N$ on a set of possible worlds s called an *information state*. This partial preorder ranks the worlds in s by relative normality, creating what he calls an *expectation pattern*. The expectation pattern $\succsim_N$ orders worlds by how well they match a set of propositions N called the *normality set* which is considered the standard of normality:

[1] Stereotypal informativeness implicatures are also anti-scalar, as they involve implicating a stronger statement than what is said (*typical road* is stronger than *road*).

(3) a) $w \gtrsim_N w' \leftrightarrow \{P \in N | P(w)\} \supseteq \{P \in N | P(w')\}$
w is *at least as normal as* w' iff the set of normal propositions true in w is a superset of the normal propositions true in w'

b) $w \succ_N w' \leftrightarrow \{P \in N | P(w)\} \supset \{P \in N | P(w')\}$
w is *strictly more normal than* w' iff the set of normal propositions true in w is a *strict* superset of the normal propositions true in w'

The example *if Tammy is a cat she should have a tail* can thus be paraphrased as follows: for every world where Tammy is a cat but Tammy does not have a tail, there is always a strictly more normal world where Tammy is a cat and Tammy has a tail.

Consider now all such expectations of normality associated with "cat" – i.e. x has a tail, x has fur, x meows, x is a pet, etc. We can treat this set of normal expectations about cats as an alternative "interpretation" of cat, its *normative interpretation*. While the usual *classical interpretation* $[\![\cdot]\!]$ returns the extension/intension of an expression in a model, the normative interpretation $\lfloor\cdot\rfloor$ returns the set of defaults associated with an expression:

(4) **a) Classical interpretation:**
$[\![\text{cat}]\!]^w = \lambda x_e[x \text{ is a cat (in } w)]$
The classical interpretation of "cat" is the set of entities in a world w which are cats by ostensive definition.

b) Normative interpretation
$\lfloor\text{cat}\rfloor = \{ \lambda x_e[x \text{ has a tail}], \lambda x_e[x \text{ has fur}], \lambda x_e[x \text{ meows}], \lambda x_e[x \text{ is a pet}] \ldots \}$
The normative interpretation of "cat" is the set of default properties of cats.

The normative interpretation of a term such as "cat" can be used as if it were a normality set to generate an expectation pattern $\gtrsim_{\lfloor\text{cat}\rfloor}$ between entities. If we say that $x \gtrsim_{\lfloor\text{cat}\rfloor} y$ for two entities x and y, then we say that x is at least as cat-like as y because the set of normal cat-like properties that x possesses is a superset of the set of cat-like properties that y possesses:

(5) a) $x \gtrsim_{\lfloor\text{cat}\rfloor} y \leftrightarrow \{P \in \lfloor\text{cat}\rfloor | P(x)\} \supseteq \{P \in \lfloor\text{cat}\rfloor | P(y)\}$
x is *at least as cat-like as* y iff the set of cat-like properties that x possesses is a superset of the cat-like properties that y possesses
b) $x \succ_{\lfloor\text{cat}\rfloor} y \leftrightarrow \{P \in \lfloor\text{cat}\rfloor | P(x)\} \supset \{P \in \lfloor\text{cat}\rfloor | P(y)\}$
x is *strictly more cat-like than* y iff the set of cat-like properties that x possesses is a *strict* superset of the cat-like properties that y possesses

In Yalcin's system, the normality set N generates an expectation pattern $\gtrsim_N$ that ranks the worlds in the information state s by general normality. Analogously, the normative interpretation $\lfloor\text{cat}\rfloor$ generates an expectation pattern $\gtrsim_{\lfloor\text{cat}\rfloor}$ that ranks the entities

in the classical interpretation $[\![\text{cat}]\!]^w$ by how typically cat-like they are. However, while the classical interpretation $[\![\text{cat}]\!]^w$ naturally falls into position as the analogue of the information state in Yalcin's system, it is important to note for the later discussion on metaphors that the expectation pattern does *not* require the comparison to be between two cats: we can easily understand that chairs are more cat-like than doors because chairs tend to have four legs.

This analysis connects naturally with the three diagnostics for defaulthood: Yalcin proposed this system to directly formalize weak necessity operators such as *should*, Thakral (2018) provides an analysis of generic quantification in terms of expectations of normality, and van Dijk (1978) describes the pragmatic component of *but* as requiring the asserted conjunction to violate an expectation of normality (though at the time the notion was not formalized).

2. Contextual suppression

Not all lexical defaults become informativeness implicatures in all contexts. For example, if I have a pet Sphynx cat (a breed that does not have fur), then mention to a friend that my cat is unwell, the informativeness implicature that my cat has fur will not arise. I propose that this is not because the implicature is generated then immediately cancelled, but rather because the default itself is *suppressed* such that the implicature is not generated in the first place.

A default can be suppressed if a) it contradicts the information already established in a discourse context (the *compatibility* criterion) or b) it fails to provide any useful information for the purpose of the conversation (the *relevance* criterion). To formulate this, I adopt a simplified model of a discourse context based on Hamblin (1973), Stalnaker (1978) and Roberts (1996). I assume that a context variable c provides (among other information not considered here) a set of propositions modelling the common ground CG and a set of sets of propositions modelling the questions under discussion QUD. The discourse context can thus be modeled as a pair $c = (CG_c, QUD_c)$.

(6) A *discourse context* is an ordered pair $c = (CG_c, QUD_c)$ such that:
- CG_c is a consistent set of propositions modelling the common ground at c – the information accepted into the conversation as of the moment of utterance
- QUD_c is the set of questions under discussion

The members of the QUD are *questions*, which are themselves modeled as sets of possible answers (*q-alternatives* in the terminology of Roberts 1996). A *full answer* to a question selects one of the possible alternative answers, while a *partial answer* reduces the number of possible alternatives.

(7) **a) Binary Question**
$[\![$Did Bill sweep the kitchen?$]\!]$
$=$ {$[\![$Bill swept the kitchen$]\!]$, $[\![$Bill did not sweep the kitchen$]\!]$}

b) Wh-Question
⟦Who ate the cake?⟧
= {⟦Ahmed ate the cake⟧, ⟦Bill ate the cake⟧, ⟦Camila ate the cake⟧ ... }

Roberts proposes a significant amount of additional discourse structure to model the development of a discourse over time, including a stack structure for QUD that allows it to distinguish between at-issue and not-at-issue content of discourse contributions. As at-issueness is not a topic I am addressing here, and because of the simplicity of the examples I am analyzing and the contexts in which they are embedded, the simpler set structure will suffice for my purposes, leaving the substantial question of the effects of at-issueness and discourse coherence on metaphor interpretation and informativeness implicatures more generally to future research.

Given this formulation of a discourse context, the suppression algorithm for unmarked declaratives[2] can be described as follows:

(8) **Contextual Suppression Algorithm**
Suppress any default P of an expression ω in a context c such that:
a) P contradicts any $P' \in CG_c$
b) P fails to provide a full or partial answer to any $Q \in QUD_c$
c) Else assert[3] P

The effect of contextual suppression on a normative interpretation is to prune it into a *discourse contribution*; the set of defaults associated with an expression that remain unsuppressed given a context c. I write the discourse contribution of an expression ω in a context c as $\mathbf{[}\omega\mathbf{]}^c$.

2.1 Example calculations with sentences interpreted literally

I now illustrate the mechanism in action generating informativeness implicatures with sentences interpreted literally. With literal interpretations, we assume that the classical interpretation is contributed to the discourse and consider only contextual suppression of defaults and the generation of supplemental informativeness implicatures. In order to abstract over the substantial issue of compositionality among the Yalcin-esque lexical representations I introduced in §1, I consider in the examples below the algorithm as it applies to a single lexical item as the core of a nominal predicate.

[2] Focus or other specialized discourse markers that I am not considering may be employed to alter this process.

[3] As I am not addressing at-issueness, I use "assert" loosely here; I do not intend to suggest that informativeness implicatures are necessarily at-issue.

(9) Context: Andy and Sanna are talking about a "Tammy" who Sina does not know.
Sina: Who is Tammy?
Sanna: Tammy is a **cat**.

a) CG contains no information about Tammy common to all participants
b) QUD accepts any information about Tammy
c) $[\![\text{Tammy is a cat}]\!]^w$ is contributed by assumption
d) $\mathbf{[}\text{cat}\mathbf{]} =$
- x has a tail
- x has fur
- x meows
- x is a pet
- x has good balance
- x survives falls from great heights

e) $\mathbf{[}\text{Tammy is a cat}\mathbf{]}^c =$
- Tammy has a tail
- Tammy has fur
- Tammy meows
- Tammy is a pet
- Tammy has good balance
- Tammy survives falls from great heights

No default of *cat* is suppressed; all are asserted as informativeness implicatures in this context. Given that this is the first information that Sina is receiving about Tammy, he employs all of the available defaults to enrich his understanding of the topic of the conversation. This is still the case if Andy and Sanna know that Tammy is a Sphynx cat; given that Sina does not know that Tammy is a cat at all, if he is simply told that Tammy is a cat he will presume that Tammy has fur.

(11) Context: Andy and Sanna are talking on Sanna's first floor apartment balcony, with Tammy (Sanna's pet Sphynx cat) walking on the railing.
Andy: Won't she fall?
Sanna: [Tammy is a **cat**], she will be fine.

a) CG contains much information about Tammy, including how she is atypical for not having fur.
b) QUD accepts only information relevant to Tammy potentially falling off the balcony.
c) $[\![\text{Tammy is a cat}]\!]^w$ is contributed by assumption
d) $\mathbf{[}\text{cat}\mathbf{]} =$
- x has a tail
- x has fur
- x meows

- x is a pet
- x has good balance
- x survives falls from great heights

e) ⟦Tammy is a cat⟧^{c} =

- ~~Tammy has a tail~~ (suppressed due to relevance violation)
- ~~Tammy has fur~~ (suppressed due to compatibility violation)
- ~~Tammy meows~~ (suppressed due to relevance violation)
- ~~Tammy is a pet~~ (suppressed due to relevance violation)
- Tammy has good balance
- Tammy survives falls from great heights

Sanna utters "Tammy is a cat" in this context in order to convey the informativeness implicatures that she will not fall because she has good balance, and that even if she does fall she will survive because cats can survive falls like that. Other default properties of cats relating to their appearance or habits are suppressed given information already available to both interlocutors, and thus do not arise as informativeness implicatures in this context.

3. Metaphor interpretation

Metaphor is often pushed out of the domain of a theory of meaning entirely. For example, Nunberg (1977) posits that metaphor generates modified interpretations, but the precise mechanism behind these modified interpretations are not of interest to a theory of meaning; Davidson (1979) proposes that it is a mistake to call the content conveyed by metaphors a "meaning" at all, and Lapore and Stone (2010) agree, concurring that as metaphor does not commit a speaker to specific truth conditions, it is therefore meaningless.

Davidson (1979) argues that a metaphor, much like a picture, can "make us appreciate some fact – but not by standing for, or expressing, the fact" (46). He points out that there is no end to what a metaphor can call to attention, and attempting to express that content as propositional information leads to an undecidable theory of meaning. He does not deny that metaphor can have an effect of discourse, but instead he denies that this effect stems from metaphor having a meaning to be grasped. Lapore and Stone (2010) elaborate on Davidson, arguing that there is more to gain from distinguishing metaphorical content and meaning then to try to give them a unified treatment, reducing meaning to truth conditional content and metaphorical content to a post-hoc creative process.

Nunberg (1977) comes to a similar conclusion to Davidson in his case study of the passage "The castled crag of Drachenfels *frowns* o'er the wide and winding Rhine" from Lord Byron's poem "Childe Harold's Pilgrimage." He explores various spatial relations and the emotive personification of the landscape that the use of the term *frowns* evokes, concluding that the inferences to be drawn from the term are endless and indeterminate. Unlike Davidson, however, his analysis posits that there is a proposition such as $frown^*(crag)$ expressed here, where $frown^*$ is a predicate contextually modified from $frown$, though he argues that determining how this modification is carried out is not in the scope of a theory of meaning.

I do not adopt the position here that the meaning-like content of figurative language and meaning proper are distinct, or that the explication of figurative content falls outside the scope of a theory of meaning. A theory of figurative meaning does not need to be an exact and complete description of the process of figurative interpretation to provide insight, nor does it need to be entirely computable to achieve explanatory adequacy and be empirically productive.[4] Even if Davidson, Lapore, and Stone are correct in positing that the content of figurative language is not entirely (or at all) propositional in nature, the content can still be approximated by a (possibly infinite) set of propositions. If a theory of meaning can then specify operations that correctly predict the behaviour of finite subsets of these propositional approximations and explain/predict the contribution of figurative language to wider discourse structures, then it is a fruitful theory of figurative meaning.

In contrast, some important works that have made headway on a formal analysis of metaphor interpretation include Grice (1975), Searle (1979), Hobbs (1990), Hills (1997), and Camp (2003). Grice (1975) briefly treats metaphor as a product of repair when an assertion is categorically false; we thus search for a metaphorical interpretation when there is a violation of the Maxim of Quality – "Do not say what you believe to be false" (46). What these utterances then express is "some feature or features in respect of which the [subject] resembles (more or less fancifully) the mentioned substance" (53). Observing that falsehood is not a prerequisite to metaphor, Searle (1979) modifies Grice's position, treating metaphor as a product of repair when an assertion is more generally *defective* – we seek a metaphorical interpretation when an utterance exhibits "obvious falsehood, semantic nonsense, violations of the rules of speech acts, or violations of conversational principles of communication" (105).

There are two major problems with the view that metaphor arises as a product of repairing defective utterances. The first and simplest is that metaphor is ubiquitous, and a repair strategy should apply to exceptions rather than be a rule – if we have to repair more frequently than not, it seems likely that we would begin to perform this "repair" even when it is unnecessary, at which point it is simply an element of interpretation rather than something triggered by any particular property of an utterance. In fact, as Searle (1979) himself points out, we may do exactly that, and interpret utterances literally and figuratively simultaneously. Searle says that "[t]here are various other clues that we can employ to spot metaphorical utterances" (105) beyond them being defective, but the question arises again here of why we need to spot them at all when we could just run the "repair" process unprompted in the first place.

This ties into the second problem with this view, which is that there is no psycholinguistic evidence that any additional effort or processing is required to achieve metaphorical interpretation, while there is evidence that trying to suppress a metaphorical interpretation takes additional effort (e.g. Glucksberg at al. 1982), suggesting precisely

[4] Prince and Smolenski (1983) analogously argue against objections arising from computability in Optimality Theory. They argue that explanatory and descriptive adequacy are the only requirements of a well-defined theory whose objective is to constrain and evaluate the space of hypotheses, and that computability as a meta-constraint conflicts with this broader enterprise (p. 215).

what the first problem implies: that we process figurative meaning automatically, without a prompt.

The remaining authors do not presuppose that utterances with metaphorical interpretations are defective. Hobbs (1990) analyzes metaphor as the product of transferring properties from one conceptual domain to another. He identifies several kinds of inferences that are or are not carried over from an old conceptual domain into a new one in the process of metaphor interpretation, which may be present regardless of whether the utterance itself was also interpreted literally. Hills (1997) analyzes it in terms of pretense; when we read that Romeo said that "Juliet is the Sun", we make-believe that Juliet was in fact the Sun and from this act draw the inferences necessary to understand the content of Romeo's metaphor, and we may use this process to also interpret literal utterances. Camp (2003) argues similarly to Hobbs that metaphor involves a process of analogy whereby one thing is viewed in the aspect of another; if a speaker says "a is F", she intends the hearer to match the characterization of a to the most prominent features of F, which may include $F(a)$ itself.

Another domain of literature where metaphor has been discussed is Relevance Theory, where word meaning is localized and contextualized via the construction of *ad-hoc concepts* (Wilson and Carston 2007, Carston 2015; see also Rayo 2013): concepts have an encoded meaning, but this is readily modified in-situ to maximize their local relevance within a given discourse context. To relevance theorists, this process is not unique to metaphor, but is also responsible for other lexical pragmatic phenomena such as narrowing and broadening. There is thus a single ubiquitous relevance-driven algorithm that derives both "literal" and "metaphorical" uses of language.

Relevance Theory posits that the activation of a concept during the interpretation of a word evokes a bundle of associated encyclopaedic presumptive information which can then be adjusted to maximize relevance within a particular context. By selecting some of this information and filtering out the rest, the denotation and the implicatures of the word in question are *simultaneously* modified to best fit the expectation of maximum relevance within the discourse, meaning that the process spans across the division of semantics and pragmatics. In parallel with the proposals of Nunberg (1977), the interpretation of a word such as "bachelor" is considered to be a concept BACHELOR which, in the context of Billy's wife exclaiming that "Billy is a bachelor", is modified to an ad-hoc concept BACHELOR* which does not carry the entailment of being unmarried but rather various implicatures such as being carefree, fast living, and undomesticated (Carston 2015: 202).

While I do not align myself with the goal of Relevance Theory to paint general pragmatic principles as the product of a single cognitive mechanism for maximizing relevance, the particular component of the theory that pertains to metaphor interpretation aligns itself well with my proposal.[5]

3.1 Example calculations with sentences interpreted metaphorically

[5] There are, however, significant deviations between my proposal and theirs; most saliently, I do not deny that the classical interpretation is privileged despite being suppressible, and I do not claim that my approach can handle all other forms of lexical pragmatic phenomena.

The crux of my approach is that beyond generating informativeness implicatures, if we allow the contextualization mechanism to also suppress the classical interpretation, it correctly predicts metaphorical discourse contributions without requiring a repair strategy specific to metaphor interpretation.[6] To achieve this, we can no longer assume that the classical interpretation of an expression is automatically contributed; *I instead treat the classical interpretation as a member of the normative interpretation*.

Consider first an example where a literally false statement receives a metaphorical interpretation through the system otherwise responsible for deriving its informativeness implicatures:

(10) Context: Joan just witnessed a salesperson oversell a customer on a product they did not need.
Ahmed: Why are you angry?
Joan: Because [some salespeople are **snakes]**.

a) CG contains the fact that salespeople are not literally snakes, don't have scales, etc.
b) QUD accepts any information that could serve as a reason for Joan being angry.
c) ⟦snake⟧ =
- (x is a snake)
- x has scales
- x have venomous bites
- x are dangerous
- x are deceptive

d) ⟦some salespeople are snakes⟧^{c} =
- ~~(some salespeople are snakes)~~ (compatibility violation)
- ~~some salespeople have scales~~ (compatibility violation)
- ~~some salespeople have venomous bites~~ (compatibility violation)
- some salespeople are dangerous
- some salespeople are deceptive

In this context, Joan employs the stereotype of *snake* to convey the relevant properties of the salesperson's behaviour that angered her so; namely the deceptiveness in their sales tactic that makes them dangerous to do business with. The classical interpretation, blatantly stating a falsehood, is suppressed here as if it were simply a default.

[6] For this to work, contextual suppression *must* be something different from implicature cancellation – the literal semantic content of an utterance cannot be cancelled in the same way that an implicature can, so contextual suppression must be a separate more powerful mechanism.

Consider now a similar example where the system I am proposing correctly predicts that a coherent discourse contribution fails to arise, even with the availability of metaphorical interpretation:

(11) Context: Joan just witnessed a salesperson oversell a customer on a product they did not need.
Ahmed: Why are you angry?
Joan: **??** Because [some salespeople are **supercars]**.

a) CG contains the fact that salespeople are not literally supercars, don't have wheels, etc.
b) QUD accepts any information that could serve as a reason for Joan being angry.
c) **[**supercar**]** =
- (x is a supercar)
- x has wheels
- x is expensive
- x is fast

d) **[**some salespeople are supercars**]**^{c} =
- ~~(some salespeople are supercars)~~ (compatibility violation)
- ~~some salespeople have wheels~~ (compatibility violation)
- ~~some salespeople are expensive~~ (relevance violation)
- ~~some salespeople are fast~~ (relevance violation)

Joan's reply in (11) is *nonsensical* despite the availability of metaphorical interpretation. The system correctly predicts this because the salient defaults in the normative interpretation of *supercar* are *all* contextually suppressed. Because of this, Ahmed would likely be very confused by Joan's reply, and ask for elaboration. She could potentially do so by saying that "some salespeople just zoom by and don't pay any attention to their customers", retroactively making "some salespeople are fast" retrievable, but this would not reflect that she is angered by the salesperson overselling a product to an unfortunate customer.

Consider now an example where a literally *true* utterance receives a metaphorical interpretation:

(12) Context: Written on a plaque hung up on the wall in someone's house.
No man is an **island**.

a) CG contains only broad background assumptions
b) QUD is similarly broad; "what is the way things are?"
c) **[**island**]** =
- (x is an island)
- x is surrounded by water

- x is remote
- x is isolated
- x is difficult to reach

e) ⟦no man is an island⟧^{c} =

- ~~(no man is an island)~~ (relevance violation)
- ~~no man is surrounded by water~~ (relevance violation)
- no man is remote
- no man is isolated
- no man is difficult to reach

In this example, the classical interpretation as well as defaults relating to the appearance of an island are all suppressed as irrelevant because even the broad out-of-the-blue QUD is already partially answered by the same common-sense knowledge that defaults embody; the answers to the subquestion "are men islands?" is not present among the q-alternatives of "what is the way things are?" because *is not an island* is already an entailment of *man* (or at least itself a default).

Revisiting now the earlier examples with literal utterances, a surprising result arises: example (8), reanalyzed using the metaphor-enabled algorithm in (13), remains literal as the classical interpretation is conveyed; however, example (9) reanalyzed in (14) reveals itself to be a metaphor in this analysis.

(13) Context: Andy and Sanna are talking about a "Tammy" who Sina does not know.
Sina: Who is Tammy?
Sanna: Tammy is a **cat**.

a) CG contains no information about Tammy common to all participants
b) QUD accepts any information about Tammy
c) ⟦cat⟧ =

- (x is a cat)
- x has a tail
- x has fur
- x meows
- x is a pet
- x has good balance
- x survives falls from great heights

d) ⟦Tammy is a cat⟧^{c} =

- **(Tammy is a cat) (not suppressed, therefore utterance still literal)**
- Tammy has a tail
- Tammy has fur
- Tammy meows
- Tammy is a pet
- Tammy has good balance

- Tammy survives falls from great heights

(14) Context: Andy and Sanna are talking on Sanna's first floor apartment balcony, with Tammy (Sanna's pet Sphynx cat) walking on the railing.
Andy: Won't she fall?
Sanna: [Tammy is a **cat**], she will be fine.

a) CG contains much information about Tammy, including how she is atypical for not having fur.
b) QUD accepts only information relevant to Tammy potentially falling off the balcony.
c) ⟦cat⟧ =
- (x is a cat)
- x has a tail
- x has fur
- x meows
- x is a pet
- x has good balance
- x survives falls from great heights

e) ⟦Tammy is a cat⟧^{c} =
- ~~**(Tammy is a cat)**~~ **(relevance violation) !!**
- ~~Tammy has a tail~~ (relevance violation)
- ~~Tammy has fur~~ (compatibility violation)
- ~~Tammy meows~~ (relevance violation)
- ~~Tammy is a pet~~ (relevance violation)
- Tammy has good balance
- Tammy survives falls from great heights

Because Tammy being a cat is already a part of the common ground and does not provide information that addresses the QUD in the context of (14), the algorithm suppresses it, leading to the analysis that its discourse contribution is metaphorical.

4. Conclusion

I have presented some preliminary work that suggests that the interpretation of metaphors is a natural byproduct of the calculation of informativeness implicatures. I have argued that it is not possible to calculate informativeness implicatures precisely without drawing from a conventional source of lexical default information. While this source of information is non-representational, it can be approximated using expectations of normality (a la Yalcin 2016). I have then illustrated that a mechanism is necessary to suppress defaults which conflict with the information already established in a discourse to prevent them from being expressed as informativeness implicatures. Subsequently, I have demonstrated that this

mechanism produces an analysis of metaphorical interpretation when it is allowed to suppress literal meanings alongside defaults.

Much work remains to be done, such as exploring what determines which components of an utterance may be interpreted metaphorically and which must be interpreted literally, the relationship between the novel metaphors I consider and conceptual metaphors (a la Lakoff and Johnson 1980), and a more in-depth exploration of the connection between this proposal and Relevance Theory.

References

Camp, Elisabeth M. 2003. *Saying and seeing-as: the linguistic uses and cognitive effects of metaphor*. Doctoral dissertation, UC Berkeley.

Carston, Robyn. 2015. Contextual adjustment of meaning. In *the Routledge handbook of semantics*, ed. N. Riemer, 195-210. New York, NY: Routledge.

Davidson, Donald. 1979. What metaphors mean. *Critical inquiry* 5(1): 24-45.

Glucksberg, Sam, Patricia Gildea, and Howard B. Bookin. 1982. On understanding nonliteral speech: can people ignore metaphors? *Journal of verbal learning and verbal behavior* 21, 85-98.

Grice, H. P. 1975. Logic and conversation. In *Syntax and semantics, Vol. 3, Speech acts*, ed. P. Cole and J.L. Morgan, 41-58. New York, NY: Academic Press.

Hamblin, Charles L. 1973. Questions in Montague English. *Foundations of language* 10(1): 41-53.

Hills, David. 1997. Aptness and truth in verbal metaphor. *Philosophical topics* 25(1): 117-153.

Hobbs, Jerry R. 1990. *Literature and cognition*. Stanford, CA: CSLI.

Kratzer, Angelika. 1991. Modality. In *Semantik/Semantics*, ed. A. von Stechow and D. Wunderlich, 639-650. Berlin: de Gruyter Mouton.

Lakoff, George, and Mark Johnson. 1980. *Metaphors we live by.* Chicago, IL: University of Chicago Press.

Lepore, Ernest, and Matthew Stone. (2010). Against metaphorical meaning. *Topoi*, *26*(2), 165-180.

Levinson, Stephen C. 2000. *Presumptive meanings: the theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Prince, Alan, and Paul Smolensky. 1993. *Optimality theory: constraint interaction in generative grammar*. https://roa.rutgers.edu/files/537-0802/537-0802-PRINCE-0-0.PDF

Nunberg, Geoffrey D. 1977. *The pragmatics of reference*. Doctoral dissertation, the City University of New York.

Rayo, Agustín. 2013. A plea for semantic localism. *Noûs* 47(4): 647-679.

Roberts, Craige. 1996. Information structure in discourse: towards an integrated formal theory of pragmatics. *Semantics & pragmatics* 5(6): 1-69.

Searle, John. 1979. Metaphor. In *Metaphor and thought*, ed. A. Ortony, 92-123. New York, NY: Cambridge University Press.

Stalnaker, Robert C. 1978. Assertion. In *Pragmatics*, ed. P. Cole, 315-332. New York, NY: Academic Press.

Thakral, Ravi. 2018. Generics and weak necessity. *Inquiry* 1-28.

van Dijk, Teun A. 1979. Pragmatic connectives. *Journal of pragmatics* 3, 447-456.

Wilson, Dierdre, and Robyn Carston. 2007. A unitary approach to lexical pragmatics. In *Pragmatics*, ed. N. Burton-Roberts, 230-60. Basingstoke, UK: Palgrave.

Yalcin, Seth. 2016. Modalities of normality. In *Deontic modality*, ed. N. Charlow & M. Chrisman, 230-255. Oxford, UK: Oxford University Press.