Optimizing interpretation from a Generative Lexicon: a case study of Metonymic Type Coercion in modified nouns

Marlies Kluck
University of Groningen, Center of Language and Cognition Groningen
Department of Linguistics, P.O. Box 716
9700 AS Groningen, The Netherlands
m.e.kluck@rug.nl

Abstract
This paper concerns the interpretation effect that is called Metonymic Type Coercion (MTC). MTC typically involves a shift from object to image, and seems to be triggered by the combination of lexical entries. It will be argued that the current operations as proposed in GL (Pustejovsky 1995) cannot account for the shift in interpretation characteristic for MTC. I will propose an account of MTC built on GL, but integrated in the framework of Optimality Theory.

1 Introduction
Designing a model for semantics is usually accompanied by looking at the boundaries of natural language interpretation. In this paper I will discuss a phenomenon that posits an interesting challenge for compositionality-based frameworks such as the Generative Lexicon (GL). In current studies this phenomenon has been referred to as Metonymic Type Coercion (MTC). Well-studied examples of MTC are a fake gun and a stone lion (cf. Franks 1995, Kamp & Partee 1995, Coulson & Fauconnier 1999, Partee 2001), and the same effect can be seen in chocolate teapot. At first glance, it seems that some property that is normally understood in the interpretation of lion, is excluded when it is modified by the adjective stone, i.e. a stone lion cannot be a natural kind. For fake gun, the 'real-ness' of gun seems to be negated in a sense. In chocolate teapot, teapot interpreted as the shape of an object made of chocolate, which creates a difficulty for interpreting teapot in its normal function. What is common to these examples is that the object that is denoted by the noun is understood to be an image of that object, i.e. not an instance of the object itself. It is this interpretational shift from default object (and all its properties) to image that is central to this paper. I will argue that although the GL can predict coercion when the noun is a natural kind, it does not predict coercion in chocolate teapot. In addition, the coercion from natural to artifactual kind does not capture the shift to image interpretation. I will propose an account for MTC that places GL as a lexicon in the perspective of a generative mechanism for interpretation of concepts. I will do so in the framework of Optimality Theory by proposing ranked constraints of interpretation which are violable and a set of interpretation types that are based on GL but extended with the notion of prototypicality in natural language interpretation.

The paper is structured as follows: in section 2. I discuss the data along with the analysis that is provided by the operations proposed in Pustejovsky (1995, forthcoming) and describe the limits of GL in the account of MTC. Section 3. places qualia structures and types in GL in a more general perspective of conceptual interpretation, and address the properties of secondary meaning types. In section 4. I will formulate constraints on conceptual interpretation and show the process of interpretation in terms of violation of ranked constraints in OT-tableaux. Section 5. concludes this paper.

2 A GL analysis of MTC
In this section I will explore the Generative Lexicon (Pustejovsky 1995) as account for MTC. In 2.1. I will discuss variable binding and coercion in GL and describe the possible shortcomings in the present GL model for an account of MTC. In 2.2. I will extend this discussion to the limits of interpretation of GL and argue that we need to account for secondary interpretation types in general.

2.1 Failure of variable binding inside the domain of N
Informally speaking, we can define MTC as an operation that coerces the interpretation of an object to an image of that object. Although MTC can be found in various constructions, this paper is restricted to nouns that are modified by different kinds of adjectives. I will additionally consider some examples of modification by nouns
Examples (1)-(3) show MTC on the head N in the NP, which seems to be triggered by the different adjectives. When approaching this from a GL perspective, we know that adjectives define a property that should be bound to the head they modify. The nominal adjective stone in (1) selects the constitutive quale of the noun it modifies. Assuming that stone is a natural type adjective \( A_N \), and lion is \( e_N \), the derivation in (4a) would represent the semantics in GL. What this analysis fails to capture is a. that the resulting type is artifactual \( e_A \) and b. that this artifact is a representation of the object that is normally denoted by the noun. Within GL a possible solution is to regard stone as an artifactual adjective, based on the intuition that whenever stone is used as an adjective it means ‘made of stone’ (cf. Franks 1995). This certainly casts a different light on the matter; stone as artifactual type now not only looks for a concrete noun, made of stone can now be analyzed as a TELIC quale that must be bound in interpretation. For (4b), we now need coercion to derive the correct type \( e_A \) for lion. In modern GL (Pustejovsky forthcoming), this kind of type coercion would be regarded as introduction. Except for the fact that the changed ontology of lion is not explicit in this derivation, it works pretty well for (1). However, for (2), GL would not predict coercion; teapot has the selected type \( e_A \) (functional type).

The adjective fake has a similar interpretation effect in (3), namely an image of a phone, i.e. not a real phone. However, this is not the result of a change of the constitutive quale. Fake has been described in classic work on adjectives as belonging to the privative class (Kamp 1975, Kamp & Partee 1995, Franks 1995). Privatives are commonly said to have a negating property; informally this means fake N is equal to \( N \) is not an N. This has led to interesting suggestions particularly in Partee (2001) who proposes that privatives are subjective adjectives with coercion (i.e. strict privatives do not exist in natural language). From a GL perspective we can see this more clearly; the adjective does not bind the noun as a whole but selects a specific quale. Whereas in (1) and (2) the constitutive qualia are selected, fake selects the AGENTIVE (for \( e_N \)) and the TELIC quale \( (e_A) \). The semantics of fake can then be described as negating the TELIC or AGENTIVE quale of the noun. Also here, we can assume that fake has a type selection for \( e_A \). Comparable to the ‘made of’ property in stone and chocolate as adjectives, we can say that fake selects for an artificial type because its TELIC quale says ‘made as replica of’. This, however, does not lead to coercion for (3) as we would wish to predict from GL.

In (4a) I represent the derivation based on qualia selection of the adjectives. The structure in (4b) relies on the assumption that the adjectives in (1)-(3) select for \( e_A \) because of the ‘made of/as’ property in their TELIC qualia. The arrows represent the selectional features we can assume for the adjectives in the examples, where c represents the constitutive and t the TELIC quale.

1 Note that MTC is also involved in other combinations that are variations on (1)-(3): wooden horse, chocolate doorbell, fake Gucci bag, et cetera. The focus of this paper is not the types of adjectives; the aim is to show that the combination of certain lexical items can lead to MTC in the interpretation of the whole.

2 As noted by the reviewer of this paper, this class is also called the modal class of adjectives. Other adjectives belonging to this class have also been ascribed to the privative class in the mentioned literature, such as alleged. Also temporal adjectives such as ex-, former and future have this negating effect on the noun they are combined with. However, only fake results in a shift to image in the interpretation, which is not the case for the other examples in the ‘privative’ class.

(1) a stone lion
(2) a chocolate teapot
(3) a fake gun

Note that in (4a.), the meaning that is derived is ‘there is some \( x \) such that \( x \) applies to \( e^j \). This is not true for (4b.) where the adjective is assumed to select \( e_A \) for the noun they are combined with. Note that here, the interpretations that I represented do not follow strictly from GL.
2.2 Interpretation boundaries of GL

From the discussion and GL derivations above, it is clear that we are missing something in the coercion mechanism provided by GL, which accounts for MTC. Although GL can predict nouns of the natural type to shift to artifactual, a GL approach as described above does not predict for artifactual types to be coerced. Nonetheless, it is obvious that they are not interpreted inside their normal domain. The shift we are looking for is a domain-shift for the noun: from concrete entity to image of that entity. I will argue that GL forms a firm foundation for the derivations of secondary meaning types when regarded from a prototypical theory for conceptual interpretation within an optimality account for natural language interpretation. What GL lacks is a systematic account for secondary meaning types, such as the image type that is needed for MTC, but also a type for figurative interpretations. On the other hand, the account that I will propose for MTC does not exclude a possible extension of GL. Within GL, it would be possible to extend the domain of (pseudo) dot types, and assume that for any concrete entity, there is a potential image of that entity. The reason I will not do so here is based on the nature of GL; the (pseudo) dot types considered usually have a high frequency in corpora.³ For now, let us assume that MTC is an interesting phenomenon for semanticists, but as long as it is not studied by corpus research, it is better to regard it as an artifact. Finally, MTC falls outside the boundaries of type coercion mechanisms in GL, due to the fact that the meaning shift is not in the domain of lexical types, but in the domain of ontological meaning types. In other words, what I have been calling ‘secondary’ meaning types (images, metaphors) is currently not part of the generated meaning in GL.

3 Qualia structures and conceptual ontology

It is clear from what is laid out above, that we need some extension for GL to capture the ontological shift in the interpretation of MTC examples. In 3.1, I will relate Qualia to conceptual prototypes. Based on that, in 3.2 I will define interpretation types that will be part of the account for MTC.

3.1 Qualia as core properties of entities

In order to account for a shift in ontology of the object that is referred to, we need a flexible account of concepts. One of the most flexible theories of how humans process concepts is Prototype Theory (PT) as proposed in Rosch (1978). The central idea in PT is that a category can be defined as a set with a prototype. The prototype is the member that shares most properties with the other members in that category. Membership of a category can then be seen as a scalar function (for more details I refer to Kamp & Partee 1995). Here, we are not dealing with category membership, but with the degree to which the interpreted N refers to the concept N normally denotes. When approaching the lexicon from the conceptual side, the lexical structures in GL can be seen as prototypical (or default) meanings. The default meaning is interpreted as long as this does not result in a type error. This means that lion is interpreted with the assigned qualia in the lexicon as 

(4) b.

ₐN \rightarrow N

Det a

Adj stone chocolate fake

e \rightarrow \tau

N lion teapot gun

\lambda x : e_\text{A}[\text{made of } \text{Adj}(x)]

\lambda x : e_\text{A}[\text{made as replica of } (x)]

³ In Pustejovsky (forthcoming), transitions in meaning as animal/food and count/mass are regarded as pseudo-dot objects, while tree/fruit and producer/product are dot objects. This distinction seems to be based on what is referred to. That is, two different entities as in tree/fruit, or one that can switch identity, as in water count/mass. It is unclear whether the shift from object to image of that object can be defined as either one of them, since MTC seems to share properties with both pseudo and real dot objects.
here). I will use $N_{\text{def}}$ according to the definition below. Informally put, the default type interpretation refers to the prototypical core of the concept the noun denotes, and therefore the Qualia Structure (QS) is equal to the prototype that is stored in the lexicon, and the interpreted noun belongs to the extension of the set denoted by the noun.

**DEFAULT TYPE N ($N_{\text{def}}$)**

\[
\begin{aligned}
\| N \| &= N_{\text{def}} \iff \\
\| N \| &= C_p(N) :: \\
\| QS(N) \| &= QS(C_p(N)), \text{ and} \\
\| N \| &\in \left\{ \sum_{\text{ext}} C(N) \right\} 
\end{aligned}
\]

I use $N_{\text{def}}^*$ for the interpretation type that does not share all the values in the QS (not all can be interpreted in that context, for instance in *baby kangaroo*). However, the interpreted N shares at least one value for each quale. *Baby kangaroo* would have the same AGENTIVE quale, but not the CONSTITUTIVE and FORMAL qualia belong to $Q_D$, the TELIC and AGENTIVE qualia to $Q_C$. This type of interpretation of N entails that a given N can appear (diagnostic attributes) as $N_{\text{def}}$ but lacks the central qualia that are associated with it by default. In other words, if N is interpreted as $N_{\text{ima}}$, the central qualia of $N_{\text{def}}$ are not part of the interpretation of N:

**IMAGE TYPE ($N_{\text{ima}}$)**

\[
\begin{aligned}
\| N \| &= N_{\text{ima}} \iff \\
\| N \| &\in \left\{ \sum_{\text{ext}} C(N) \right\}, \text{ and} \\
\forall Q \in \| QS(N) \| : \\
\| Q \| \cap QS(C_p(N)) \neq \emptyset, \text{ and} \\
\| N \| &\neq N_{\text{def}} 
\end{aligned}
\]

Finally, assuming that every noun denoting an entity can also denote the image of that entity. I propose the following definition for *image type* interpretation. Here, I distinguish between $Q_D$ and $Q_C$, this is based on Franks (1995) where attributes are divided into central (C) and diagnostic (D). The

**CLOSE TYPE N ($N_{\text{def}}^*$)**

\[
\begin{aligned}
\| N \| &= N_{\text{def}}^* \iff \\
\| N \| &\in \left\{ \sum_{\text{ext}} C(N) \right\}, \text{ and} \\
\forall Q \in \| QS(N) \| : \\
\| Q \| \cap QS(C_p(N)) \neq \emptyset, \text{ and} \\
\| N \| &\neq N_{\text{def}} 
\end{aligned}
\]

4 The exact mechanism of Prototype Theory lies beyond the scope of this paper. For here it is sufficient to assume that any $N_{\text{def}}$ is a prototypical (default) interpretation of N, and that $N_{\text{def}}^*$ is less prototypical but still within the same domain as $N_{\text{def}}$. For more discussion about Prototype Theory, I refer to Oherson & Smith (1981) and Kamp & Partee (1995).

4 Interpretation in GL-based Optimality Theory

To complete my proposal for MTC in modified nouns, this final section concerns the incorporation of the interpretation types I described in 3., in an Optimality-theoretic approach to interpretation. It will be argued that MTC is an example of a collision of lexical meanings, such that there is no valid interpretation of the whole as far as primary interpretations of the words are involved. Optimality offers a flexible model of violable constraints, which is topic to 4.1. In 4.2, I will
4.1 Strength versus Fit constraints for conceptual interpretation in context

Combining the proposals of GL and a scale of interpretation types, a commonsensical assumption is that a default interpretation type is the preferred interpretation. This means that unless there is an error in interpretation (such as a free variable or a type error), the default type is interpreted for a given lexical entry. This is close to the fundamental idea of Optimality Theory, which has recently been explored for the domain of semantics (see Blutner et al. 2003, Hendriks & De Hoop 2001, Zwarts 2003). Especially in Zwarts (2003), it is suggested that interpretation of complex expressions is regulated by a constraint for the strongest interpretation of the elements (STRENGTH, also called FAITHFULNESS) on the one hand, and a constraint for consistent interpretation of the whole (FIT) on the other. We can now formulate two constraints for the different elements we are dealing with here, those are defined as scales with the preference for the highest interpretation type (the prototype with its QS):

**STRONG NOUN (STRONGN)**
\[ N = N_{\text{max}} \text{ on Type Scale (TS), where } TS = N_{\text{def}} \triangleright N_{\text{def}}^+ \triangleright N_{\text{imu}} \triangleright N_{\text{fig}} \]

**STRONG ADJECTIVE (STRONGADJ)**
\[ Adj = Adj_{\text{max}} \text{ on Type Scale (TS), where } TS = Adj_{\text{def}} \triangleright Adj_{\text{def}} \]

In a given AdjN construction, the strongest interpretations are preferred over weaker ones according to the type scale. Adjectives (and verbs) have variables in their structures, therefore I assume STRONGADJ is higher ranked than STRONGN. But whether two strong interpretations can be composed into one consistent interpretation of the whole is up to the FITINT constraint. Note that here, I have described consistency as ‘does not equal ⊥’. I represent the composition of two lexical entries as ‘∪’:

**FIT INTERPRETATION (FITINT)**
\[ \| QS(Adj) \cup QS(N) \| \neq ⊥ \]

This constraint is fatal when violated and thereby higher ranked than the respective STRENGTH constraints. Whenever the unification of the interpreted elements is zero, FITINT is violated and there is no consistent interpretation. For here, I assume that violation occurs when there is a free variable in the broadest sense. This can mean an unsatisfied type selection or an unbound variable that is part of the lexical structure of one of the elements, or a combination of the two.

4.2 MTC as optimal interpretation

Turning back to the examples, we can now give an analysis that accounts for MTC and the process to MTC. As shown in table 1., the types that are higher in the type scale for N (N_{def} and N_{def'}) violate FITINT, marked in table 1. with *!:

<table>
<thead>
<tr>
<th>[NpAdjN]</th>
<th>FIT INT</th>
<th>STRONG ADJ</th>
<th>STRONG N</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;Adj_{def}, N_{def}&gt;</td>
<td>*!</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>&lt;Adj_{def}, N_{def'}&gt;</td>
<td>*!</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>*&lt;Adj_{def}, N_{imu}&gt;</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>&lt;Adj_{def}, N_{def'}&gt;</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

Table 1: visualization of the interpretation process of MTC in (1)-(3)

This violation in (1) is due to either *stone* not being able to bind its variable onto N, if N is interpreted as N_{def} or N_{def'}, or an unsatisfied type selection for Adj. The first available interpretation that does not lead to a violation of FITINT is then the pair <Adj_{def}, N_{imu}>. This violates STRONGN twice (i.e. it has two candidates for interpretation that are higher on the scale), but since FITINT is higher ranked, STRONGN can be violated twice if that entails satisfaction of FITINT. In (2) *chocolate* is an adjectival use of a complex e_A noun. The CONSTITUTIVE attributes of chocolate are such that when bound to *teapot*, the TELIC quale of teapot (*‘container for boiled water’*) cannot be interpreted. This means that either there is a free variable for the meaning of *chocolate* or the noun

---

In the analysis of just AdjN examples of MTC, having three constraints seems redundant. But MTC can also occur in compounds of nouns, in which case we would have STRONGN for two elements; no ranking can be made for that. Examples of those compounds are *apple dress* and *toy truck*. I will return briefly to those in 5.
teapot has to be interpreted in a type that does not conflict with the variables that need to be bound. Interpreting N without its TELIC quale is the only option: this again is the pair \( \text{Adj}_{\text{def}}, N_{\text{ima}} \). However, chocolate is different from stone. Being derived from a complex artificial noun, it does not only bind to the CONSTITUTIVE but also to the TELIC quale to N. So, teapot in (2) is an image, but the whole construction is a chocolate. Finally, fake can bind its property onto both \( e_x \) and \( e_y \), but since it denies exactly the quale that are subsumed in \( Q_x \), the only candidate interpretation that satisfies \( \text{FTIINT} = \text{Adj}_{\text{def}}, N_{\text{ima}} \). A last question that arises here is whether default interpretation of the nouns that in MTC shift to image interpretation, follows correctly from this account. Elaborating on (2) of this paper, we can also modify teapot with earl grey (N) or paper (denominal Adj). In earl grey teapot, earl grey can only describe the TELIC quale of teapot. The QS of earl grey will contain at least that it is a kind of tea. There is nothing in its QS that cannot be bound in the interpretation of the whole. Since \( N_{\text{def}} \) is a higher ranked interpretation type, \( N_{\text{def}} \) will be preferred as long as \( \text{FTIINT} \) is not violated. The latter is not the case in earl grey teapot; both teapot and earl grey can be interpreted as \( N_{\text{def}} \) in this combination. This does not hold for paper teapot. Paper describes the CONSTITUTIVE quale of teapot in this example. If teapot is interpreted as \( N_{\text{def}}, \) paper cannot bind its variable. Paper can only be interpreted if the TELIC quale that is part of teapot (\( N_{\text{def}} \)) is eliminated from the interpretation. In other words, \( N_{\text{ima}} \) is the optimal interpretation for paper teapot, as it is for chocolate teapot.

5 Conclusion

This is a first step towards a GL-based model for interpretation that accounts for secondary meaning types. In order to capture the interpretation of MTC, I have formulated interpretation types based on GL qualia structures. Within the adopted model of OT, the result is that the default interpretation type relies fully of what GL would predict, but in when the composition of a modified noun leads to a violation of \( \text{FTIINT} \), the additional meaning types account for secondary meaning types that are available in the human interpretation mechanism. The model as presented here ends with \( N_{\text{ima}} \), but the model is extendable to capture figurative interpretation. Even more so, it is essential to formulate this type to avoid over-generation of \( N_{\text{ima}} \) interpretations for nouns in contexts where a figurative interpretation is the desired type for a given noun (\( N_{\text{fig}} \)).

As mentioned, MTC is not restricted to Adj-N constructions. It is also visible in compounds, for example toy truck and apple dress. The interesting difference between those and the examples in this paper is that compounds of this type can shift in both directions. For apple dress the most salient interpretation would be a dress with apples depicted on it, while a toy truck is a toy in the shape of a truck. When nouns are arguments of a verb, MTC can occur as a way to satisfy that requirement, like in Tom tripped over a house. Unless Tom is known to be a giant, house will be interpreted as a small replica of a house (like a doll house). What nicely follows from this model is the MTC like interpretation of examples that are the reverse of the ones discussed here, like talkative tree or affectionate chair, i.e. examples in which the adjective selects a noun with a quale for animacy, which would only be a subset of natural kinds. The satisfaction of this animacy requirement leads coerces \( N_{\text{ima}} \) when the realized N does not have that quale itself. This is without coercion to a natural kind, which is impossible but would be predicted from standard GL. The fantasy-like interpretations for tree and chair in these examples fit the definition of \( N_{\text{ima}} \) and follow from the same constraints as formulated above.

6 Acknowledgements

This research was financially supported by the Netherlands Organisation for Scientific Research (NWO). I thank Henriëtte de Swart, Joost Zwarts, Mark de Vries, Petra Hendriks, Rick Nouwen, Jennifer Spenader, Anja Lobanova and Herman Heringa for their useful suggestions for this part of my research. In addition, I wish to thank the anonymous GL2007 reviewer for the suggestions that led to the present version of this paper.

References


Coulson, Seana and Gilles Fauconnier (1999), ‘Fake Guns and Stone Lions: Conceptual
Blending and Privative Adjectives’. In Barbara Fox, Dan Jurafsky & Laura Michaelis (eds.), Cognition and Function in Language, Palo Alto, CA. CSLI.


