Grammaticality and Meaning Shift

Márta Abrusán[1], Nicholas Asher[2] and Tim Van de Cruys[2]
1. Institut Jean Nicod, CNRS, ENS, EHESS, PSL Research University, Paris, France
2. Institut de Recherche en Informatique de Toulouse, CNRS, Toulouse, France

pre-print, 2018
Final version to appear in G. Sagi et J. Woods (eds.), The Semantic Conception of Logic. Cambridge University Press

1 Introduction

Acceptable sentences are all alike; every unacceptable sentence is unacceptable in its own way. Observe the following examples:

(1) a. *There is every fly in my soup. (Barwise and Cooper 1981)
b. *Some students but John passed the exam. (von Fintel 1993)
c. *Mary is taller than no student is. (Gajewski 2008)
d. *There are any cookies left. (Chierchia 2013)
e. *How fast didn't you drive? (Fox and Hackl 2007)
f. *How tall do you regret that you are? (Abrusán 2007, 2014)

For each of the above sentences it has been argued that their unacceptability follows once we recognise that their semantics expresses something logically false or true. Yet, if logical triviality leads to ungrammaticality, then what distinguishes the examples above from the examples in (2), which are acceptable or at worst semantically anomalous?

(2) a. Every woman is a woman.
b. This table is red and not red.

An answer to this question was proposed in Gajewski (2002). According to Gajewski, there is a formally definable subset of trivial sentences, namely L-trivial sentences, whose members are systematically ungrammatical. L-triviality is calculated in the following way: replace non-uniformly all the non-logical vocabulary by a fresh variable in the logical form of a sentence S. If the resulting representation is trivial, then S is L-trivial.

This proposal has been quite influential in the literature (cf. Fox and Hackl 2007, Chierchia 2013, Abrusán 2007, 2014, among others). However, it has also been observed that the theory needs to accommodate a number of non-trivial restrictions if it is to be applied to explain data as in (1) (cf. Abrusán 2014). In particular, on closer examination of these types of data, the calculation of L-triviality needs to be supplemented with a number of ad-hoc assumptions in order to restrict the replacement procedure in one way or another. Another problem we think is the radical modularity assumption that follows from Gajewski's proposal. Grammar, viewed as a logical deductive system, is encoded in functional/logical vocabulary and is blind to the content of lexical words. This suggests that grammar is insulated not only from conceptual systems, general world knowledge, but also from most of the information encoded in lexical items. One consequence of this view is that we need to be able to distinguish logical from non-logical vocabulary on principled grounds and thus the logical and conceptual aspects of meaning need to map into two distinct types of vocabulary: the logical vocabulary and the lexical vocabulary. This consequence, as it was noted by Gajewski himself, is non-trivial. A second consequence, emphasised in Del Pinal (2017), is that a deductive system that operates on logical skeletons is a rather exotic system for which most classical formulas and rules of inference are invalid. As Del Pinal (2017) argued, this is problematic for some of the accounts that are based on the idea of L-triviality. Finally, the idea that the judgments of ungrammaticality in (1) are due to problems of *evaluations* of logical form seems conceptually wrong; it's not that these sentences are false in all models consistent with the meanings of logical words or trivial; they are word salad. They aren't even evaluable.

So although we believe a semantic explanation of the judgments in (1) is the right way to go, we don't believe that this has to do with isolating a particular kind of logical form or a class of models that such logical forms would yield. It has rather to do with semantic constraints that must be met, in order for a logical form to be constructed. More particularly, we suggest that the examples in (1) reflect a compositionality problem; in this sense they are similar to well-known examples of semantic anomaly, e.g. (3):

(3) a. #Tigers are Zermelo-Frankel sets.
b. #Colorless green ideas sleep furiously.

Following Asher (2011), we propose that a semantic anomaly is the result of a type presupposition that cannot be satisfied.

This suggestion immediately gives rise to two questions: Firstly, not all type conflicts lead to unacceptability. What is the difference between resolvable type conflicts and unresolvable ones? Second, there is an intuitive difference between classic examples of semantic anomaly such as the ones in (3) and the seemingly ungrammatical examples in (1). What explains this difference?

The process of integrating a lexical meaning into an interpretive context, can also shift a lexical meaning or introduce new meaning components as a result of the integration, depending on the type presuppositions of predicates and their arguments. Coercions are an example of a construction where type conflicts can be resolved, because the predicate licenses the addition of material needed to accommodate the type presuppositions of the argument. To a limited degree, we can also resolve type conflicts in the case of semantic anomaly. To do so, we need to apply meaning shift to relax the type presuppositions in question. Depending on the semantic distance between the actual and the target types, type shift allows us to make sense of examples of semantic anomaly to a greater or smaller degree.

(4) a. #Tigers are human.
b. #Tigers are teabags.
c. #Tigers are Zermelo-Frankel sets.

The higher the type that is involved in the type clash, the more difficult it is to shift the meaning of predicates to resolve type conflict. With very high types, meaning shifts will not be able to help. Both the classic examples of semantic anomaly in (3) and the examples in (1) involve high-level types.

In order to get a better understanding of the nature of meaning shift involved in semantic composition, as well as the restrictions on meaning shift alluded to above, we turn to a computational method called distributional semantics (DS). This computational method can find low-level types and capture corresponding meaning shifts. It does not, however, predict shifts with high-level types because either (a) the types in question denote context-invariant logical meaning that is simply invisible for distributional methods or (b) the conflicting type-clash is attached to a type so high in the type hierarchy that the type has no neighbours that share the same syntactic/semantic dependencies and so there is nowhere in the space for its meaning to shift. The fist case corresponds to examples in (1), the second case to classic examples of semantic anomaly such as (3).

Another aspect that might underlie the difference between semantic anomaly and (1) is the locality of the type-conflict. In the case of classic examples of semantic anomaly, type conflict (presupposition failure) arises at the level of predicate-argument composition. If shifting were to occur, it would also happen at this level, where the nature of the type-conflict is clear and lowest common types are easy to calculate. In the case of examples in (1), the type conflict arises at a more global level with more linguistic elements (and types, as a consequence) that are involved. Calculating lowest common types is harder, and it is more likely there are simply none.

Our proposal assumes that logical and lexical aspects of meaning do not map neatly to two different types of words (see also Abrusán et al. 2018). There is no purely logical vocabulary, and purely lexical/conceptual vocabulary. Instead, this distinction cross-cuts word boundaries; all words have both aspects of meaning. In particular, the meanings of lexical as well as logical words have both logical aspects (their model theoretic meaning) and lexical/conceptual aspects. We cannot neatly separate grammar and conceptual knowledge because they are packaged together within lexical entries. However, we can distinguish conceptual content that is contextually invariant from shiftable conceptual content. Conceptual content that supports logically valid inferences for first order definable quantifiers (whose conceptual content in the form of proof rules can determine all logical consequences of such quantifiers) should always be contextually invariant, since a particular context should not render logically valid inferences incorrect. We would never expect meaning components that lead to logically valid inferences to shift in context. Other types of conceptual content of logical words, however, can be shiftable; that is, context might affect the conceptual content and the extension of a predicate or its argument as well.

2 Logical forms, skeletons, triviality and ungrammaticality

In the analytic tradition, logical form is assumed to be a translation of the meaning of a sentence of natural language into a more ideal language that allows to calculate inferences of the original sentence more perspicuously than the natural language sentence itself. Under this conception, logical relations of the original sentence can be formally explained if they can be deduced from a formal principle that applies to the logical form of the sentence involved. In the tradition of Chomskyan generative grammar, logical form is a syntactic/semantic representation that registers the effect of certain operations that do not have a phonological effect, e.g. quantifier raising. On this conception, logical form is a representation that encodes all semantically significant features of a sentence. On both views, logical form is a level of representation where ambiguities have been resolved. It is also a representation that allows the calculation of being a logical truth or logical validity. By humans, familiar with the rules of logic.

In an influential paper, Gajewski proposed that a certain type of logical triviality has consequences for grammaticality (cf. Gajewski 2002). In particular, he proposed that there is a formally specifiable subset of trivial sentences, that he calls ‘L-trivial’, whose members are systematically unacceptable.[1] On the standard, Tarskian conception, logical truth and consequence are defined in terms of variation of truth across all interpretations of the logical form. L-triviality, in contrast, is calculated on modified logical forms in which all lexical material has been ‘bleached’, i.e. replaced non-uniformly by a fresh variable of the appropriate type. Gajewski (2002) calls these impoverished logical forms ‘logical skeletons’:[2]

(5) Logical skeletons are obtained from the logical form α as follows:

a. Identify the maximal constituents of α containing no logical items.

b. Replace each such constituent with a distinct variable of the same type.

[1] Gajewski’s (2002) original definition is about L-analyticity, but this might be slightly misleading, hence we use L-triviality, similarly to some other authors, e.g. Del Pinal (2017).

[2] Gajewski (2002) works with the notion of logical form that arises from generative grammar, cf. Heim and Kratzer (1998).

Thus the logical form of (6) is something akin to (6-a) (where *raining* stands for the proposition that it is raining), but its logical skeleton is (6-b), where p and q stand for propositional variables:

(6) It is raining or it is not raining.
 a. *raining* $\vee\neg$*raining*
 b. p or not q

Given logical skeletons, L-triviality can be defined as follows:

(7) An LF constituent a of type t is L-trivial iff a's logical skeleton receives the denotation 1 (or 0) under all interpretations.

L-triviality and unacceptability are linked in the following way:

(8) A sentence is ungrammatical if its Logical Form contains a L-trivial constituent.

Under the resulting picture, grammar itself is endowed with the capacity to calculate L-triviality. This means that the grammar of natural languages has to include (or at least interact with) a system of 'natural logic', or a 'natural deductive system' (see Fox and Hackl 2007, Chierchia 2013). By assumption, this deductive system is blind to conceptual information and cannot 'see' non-logical terms, it operates only on the basis of functional terms. This in turn presupposes that terms can be sorted into two non-overlapping classes, lexical terms and functional (or logical) terms.

In the remainder of this section we first discuss two examples of applications of L-triviality. Second, we go on observing some problems that motivate us to look for an alternative explanation.

2.1 Two examples

Exceptive constructions In his 1993 paper, von Fintel proposed that the restricted distribution of connected exceptive phrases (e.g. *but John*) can also be explained by appeal to triviality. Exceptives formed with *but* are compatible with positive and negative universal quantifiers (*every, no, none, all,* etc.), and incompatible with any other quantifier:

(9) a. *Every/No* boy but John smokes.
 b. **Some/*three/*many/*most/*less than three* boys but John smoke.

The semantics that von Fintel (1993) gives for exceptive *but* assumes that the argument of *but* is the least (i.e. the unique minimal set) one has to take out of the restrictor to make the statement true:

(10) $[\![\text{but}]\!]$=(C)(A)(D)(P)=1 iff
 $C \neq \emptyset$ and D(A-C)(P)=1 and $\forall$S [D(A-S)(P)=1 $\rightarrow$ C $\subseteq$ S]

According to the above, *Every boy but John smokes* means that C=John is the unique minimal set one has to take out of the domain A=⟦boy⟧ of the quantifier D=⟦every⟧ to make the statement *Every boy smokes* true. Why is it that only universal quantifiers can host *but*-exceptives? The problem is that with (almost) all other quantifiers, such least-exceptions lead to a contradiction.

Universal quantifiers are left downward monotone, while existential quantifiers such as *some, a, (at least) one, two,* are left upward monotone.[3] In his article von Fintel (1993) shows that modifying left upward monotone quantifiers with a *but*-exceptive always leads to a contradiction.

(11) A determiner D is left upward monotone if
for all models M = $< E, [\![.]\!] >$, and all $A \subseteq B \subseteq E$,
if $X \in [\![D]\!](A)$ then $X \in [\![D]\!](B)$.

Intuitively, left upward monotonicity captures the inference from sets to supersets on the left argument of the quantifier. For quantifiers that are upward monotone on their restrictor argument, it is always the case that if the statement D(A-C)(P) is true, then D(A)(P) is true as well, thus one could have always taken out less than C (in our case, the empty set) from A to make the statement true. But this means, that the second clause in (10) (namely $\forall$S [D(A-S)(P)=1 $\rightarrow$ C $\subseteq$ S, that requires that every alternative subset of A that one could have taken out of A to make the statement true has to be a superset of C) is false in the case of left upward monotonic quantifiers. Therefore, modifying such quantifiers with a *but*-exceptive results in ungrammaticality. Left downward monotone quantifiers, however, do not lead to a problem.

Gajewski (2002, 2008) shows that the problem persists once the LF of an exceptive sentence is transformed into a logical skeleton:

(12) a. Some boy but John smokes
b. Logical skeleton of (12-a): some [P_1 but P_2] P_3
c. Interpretation : $[\![\text{some}]\!](I(P_1)\text{-}I(P_2))\ (I(P_3))=1$
and $\forall S\ ([\![\text{some}]\!](I(P_1)\text{-}I(P_2))\ (I(P_3))=1 \rightarrow I(P_2) \subseteq S)$

Presumably, the connective *but* and the quantifier *some* are functional words and are not replaced in the logical skeleton (Gajewski uses invariance as a criterion, we come back to this shortly). Because of the left upward monotonicity of the quantifier *some*, whatever the interpretation of P_1 and P_2, if the sentence is true with ($I(P_1)$- $I(P_2)$) as the domain of the quantifier *some*, it will also be true with ($I(P_1)$) as its domain. For this reason, the second clause in the interpretation will always be false, at least as long as P_2 is prevented from being empty. Therefore all interpretations of (12) will map it to false and it is L-trivial, hence it will be predicted to be ungrammatical.

In contrast, observe again the case of simple contradiction:

[3]One problem for the theory is the quantifier *most*, as noted by von Fintel (1993) himself. See also Gajewski (2002) for further problems.

(13) This table is red and not red

  a. Logical skeleton: [this P_1 is P_2 and not P_3]

In this case, once we remove the identity of the non-logical expressions to create the logical skeleton, we cannot deduce triviality anymore. The algorithm for forming logical skeletons assigns distinct variables to the two occurrences of *red*, P_2 and P_3. Clearly the resulting logical skeleton is not L-trivial. If the interpretation of P_2 is not the same set as the interpretation of P_3, triviality does not follow.

Weak islands Abrusán (2007, 2014) discuss the problem of weak islands. Weak islands are contexts that are transparent to some but not all operator-variable dependencies. Some paradigmatic cases of weak island violations include examples of degree extraction in (14-a) and (15-a), as opposed to the acceptable questions about individuals in (14-b) and (15-b):

(14) a. *How tall isn't John?
  b. Who didn't John invite?

(15) a. *How tall do you regret that you are?
  b. Who does John regret that he invited to the party?

The traditional analysis derives these facts from a syntactic contrast (cf. Rizzi 1990 and much subsequent work). Szabolcsi and Zwarts (1993) noted, however, that the oddness of the a. examples above has a semantic flavour: they simply do not seem to make much sense. More recently, Kuno and Takami (1997) and Fox and Hackl (2007) showed that many weak island violations can be ameliorated adding certain quantificational elements such as modals to the sentence, cf. ecample (16). The fact that modals can obviate weak islands is unexpected and hard to explain on a syntactic analysis.

(16) How much wine are you not allowed to drink?

Abrusán (2007, 2014) put forth a novel semantic theory that could explain all the basic examples of weak island violations as well as the cases of modal obviation. The central thesis of this work is that these islands arise because they are predicted to lead to a contradiction at some level. There are two ways in which a contradiction can arise. In the case of factive islands, the question always has a contradictory presupposition. Observe first that presuppositions project universally from question alternatives: (17) presupposes that you invited all these ten people:

(17) Who among these ten people do you regret that you invited?
  presupposes: you invited these ten people

Universal projection is not problematic in the case of questions about individuals because the answers are independent. But in the case of degree and manner questions it leads to problems. Assume, following Schwarzschild and Wilkinson

(2002), that degree predicates relate individuals and intervals of degrees.[4] Then the presupposition of (18) is that your height is contained in every interval on some scale.

(18) *How tall do you regret that you are?

a. 'For what interval I, you regret that your height is in I?'

b. $[\![Q]\!]^{C,w}$={λw. you regret that your height∈I in w| I∈D_I}

(19) Presupposition of (18): ∀I∈D_I: your height ∈I
'your height is contained in every interval in D_I'

This presupposition is contradictory and cannot be met in any context: assume I_1 and I_2 are two non-overlapping intervals on the scale. Since heights are points on a scale, they cannot be contained in both of these intervals.

In the case of negative and wh-islands a contradiction arises in a different manner. We illustrate here a negative degree island. Following Dayal (1996) and Fox and Hackl (2007), Abrusán (2007, 2014) assumed that questions presuppose that they have a unique maximally informative answer, i.e. a true answer that logically entails all the other true answers.

(20) *Dayal's (1996) presupposition (aka Maximal Informativity Principle (MIP))*
Any question presupposes that it has a maximally informative answer, i.e. a true answer which logically entails all the other true answers.

If the MIP cannot be met, the statement for any potential answer that it is the maximally informative answer to some question is bound to state a contradiction.

In the case of negative and wh-islands the MIP cannot be met and therefore the exhaustification of any answer *Ans* to a negative degree question *Q* expresses a contradiction.[5] Observe (21) (as above, degree predicates are assumed to range over intervals):

(21) *How tall isn't John?

a. 'For what interval I, John's height is not in I?'

b. $[\![Q]\!]^{C,w}$={λw. John's height is not ∈I in w| I∈D_I}

If a degree *d* is not contained in an interval *I*, it follows that it is also not contained

---

[4]According to this proposal, the denotation of a degree (interval) predicate such as *tall* is as follows:

(i) $[\![\text{tall}]\!] = \lambda I_{\langle d,t\rangle}$: I is an interval. λx.x's height ∈I

[5]Exhaustification of answers, following Fox and Hackl (2007), is defined as follows:

(i) $Exh(Q)(Ans)$ is true iff

a. *Ans* is true

b. for any alternative proposition φ in *Q*, if φ is true, then φ is entailed by *Ans*.

in any subinterval of *I* and nothing follows wrt. to *d* for any other interval in the domain. Given this entailment pattern, an exhaustive answer to (21) amounts to the following:

(22) Exh (Q)(John's height is not $\in I_1$) is true iff

a. John's height is not $\in I_1$

b. For every interval $I_2 \in D_I$, if John's height is not in I_2, then I_2 is contained in I_1.

If John has a non-zero height then there is no answer to (21) that can be exhaustified as in (22) without leading to a contradiction.[6]

Examining the nature of the contradiction that arises in the cases of weak island violations one might wonder whether it is an instance of L-triviality of Gajewski (2002). The logical form of the exhaustive answer is as follows:[7]

(23) Exh ({John is not I-tall | $I \in D_I$})(John is not I_1-tall)

From this, the logical skeleton of an exhaustive answer to (21) can be obtained by replacing *John* and the predicate *tall* with fresh variables a_i and P_i respectively:

(24) Exh ({ a_i is not P_i | $i \in \{1,...,n\}$ })(a_1 is not P_1)

Strictly speaking, this logical skeleton is not contradictory. For it to be trivial, we have to assume that (a) the values for a_i and P_i are held constant across the question alternatives, (b) the restriction that degree predicates operate on degree scales and that they range over intervals is maintained when creating the logical skeleton.[8]

Note that these restrictions are not unique to Abrusán's account. Similar restrictions need to be assumed by Gajewski (2008), Fox and Hackl (2007), Chierchia (2013) — see Abrusán's (2014) Chapter 6 for further discussion on this. But such ad-hoc assumptions are hard to accept if we are looking for a principled account of ungrammaticality based on L-triviality.

2.2 Problems with L-triviality

Restrictions As we have seen above, to maintain that certain ungrammatical logical truths and falsities are indeed L-trivial in Gajewski's sense, we need to make a number of ad-hoc assumptions, some relatively innocent, others more problematic. We need to assume that grammar contains various restrictions on the domains

[6] If John's height is zero, then the most informative answer, the proposition that John's height is not included in (0; +∞), must be already already entailed by the common ground. Abrusán and Spector (2010) argue that a maximally informative answer must also be contextually informative.

[7] NB: In the examples above the interpretation of the relevant propositions/questions was represented for better readability, not their logical form.

[8] In the case of factive islands one more important assumption is needed to make these examples L-trivial, namely that the factivity of factive predicates is preserved when creating the logical skeleton.

of predicates (cf. Gajewski 2008, Chierchia 2013, and Abrusán 2014). Gajewski (2008) suggests that these might be represented as semantic presuppositions, in which case failure to meet them results in undefinedness. Accordingly, the definition of L-triviality is changed as follows:[9]

(25) A sentence S is L-trivial iff S's logical skeleton receives the truth value 1 (or 0) on all interpretations <u>in which it is defined</u>.

We further need to assume that the process of creating logical skeletons interacts with the process of alternative generation in a special way, namely it has to use the same constants/variables in each alternative to replace non-logical words, except the focused/questioned word (Fox and Hackl 2007, Chierchia 2013, and Abrusán 2014). Even more problematically, as discussed in Abrusán (2014), lexical presuppositions would have to interact with the mechanism checking L-triviality in a rather ad hoc fashion.

Defining logical words Gajewski's (2002) proposal hinges on distinguishing two types of vocabulary, logical and non-logical vocabulary. Finding a conceptually motivated account for this division is one of the major long-standing issues in the philosophy of logic. The most well-known account is due to Tarski (1986) who defined the difference between logical and non-logical words in terms of permutation invariance (see also Sher 1991, van Benthem 1989, 2002, Bonnay 2006, 2008, among others). The intuition behind this approach is that invariant elements do not depend on the identity of the particular individuals in the domain. Gajewski (2002) also follows this tradition, more precisely van Benthem's (1989) extension of the permutation-invariance idea to typed languages.

Permutation invariance as a definition for logical items is a relatively simple idea, though not without any problems, both from the logical and the empirical perspective (see for example van Benthem 2002 and Bonnay 2006 for an overview). However, from the linguistic perspective, it seems to include both too many and too few items. It includes too many, because, as shown by Gajewski (2009), it predicts predicates such as *self-identical* and *exist* to be logical, although we have the sense that they are not. On the other hand, as mentioned by van Benthem (2002), it excludes items in natural language that intuitively should count as logical: for example the quantifiers *every* and *each* in natural language carry the restriction that they can only quantify over countable objects, hence the sentences **Every salt is on the table, *Each milk is in the fridge* are unacceptable (unless the domains of salt and milk have been somehow individuated in the context). In contrast, the quantifier *all* can combine with mass nouns as well: *All the salt is on the table, All the milk is in the fridge*. The sensitivity of some quantifiers to the countability of the predicates they combine with makes them not permutation-invariant, hence not logical on the permutation invariance theory. Many alternative versions of the

[9] Alternatively, we need to postulate a special module of grammar, DS that enforces these restrictions and at which ungrammaticality is calculated (see Fox and Hackl 2007).

basic invariance idea exist, which characterise logicality as invariance under some other transformation (for example Feferman 1999 (relation-invariance), Bonnay 2008 (invariance under potential isomorphism), etc.). There are also many conceptually different accounts of logicality, for example proof-theoretic or algorithmic accounts (see van Benthem 2002 and references therein), or accounts that extract logical items from consequences (Bonnay and Westerstahl 2012). Neither of these are problem-free however, or significantly better suited for the linguistic purpose that we are concerned with. An alternative possibility, one that Gajewski also considers, is to replace the logical/non-logical distinction with the functional-lexical distinction familiar from the linguistic literature (cf. Abney 1987, von Fintel 1995). This too however suffers from difficulties, as some words, for example prepositions or the word *there* are not clear cases of either category.

It seems that for the moment there is no foolproof method that can distinguish logical words from non-logical ones that also makes the cut in a linguistically intuitive way. It cannot be excluded that such a property could be found in the future. But at least, the difficulties mentioned above suggest that the logical/grammatical aspects and the conceptual aspects of meaning do not map neatly onto two different classes of words. Instead, both functional and lexical words might have logical and conceptual aspects of meaning, packaged together.

Deductive system The conception of grammar and a natural deductive system that follows from Gajewski's (2002) proposal has profound implications for how we should think about the language system and its interaction with other cognitive systems in general. First, it suggests a very radical form of modularity of language: grammar is insulated not only from conceptual systems, general world knowledge, but also from most of the information encoded in lexical items. Second, if L-triviality can have implications for grammaticality, the grammar (or the language module, language organ, or whatever) needs to contain — or at least interact with — a natural deductive system (cf. Fox and Hackl 2007).

As it was emphasised in Del Pinal (2017), a deductive system that operates on logical skeletons is a rather exotic system for which most classical formulas and rules of inference are invalid. It is conceivable that the properties of the natural deductive systems, as used by grammar, could be radically different from classical systems. However, Del Pinal (2017) argues that certain key accounts depending on logical skeletons, e.g. Chierchia's (2013) account of polarity-sensitive items, cannot be maintained if the Law of Non-Contradiction is invalid at the level of representation where the deductive system determines grammaticality.

In addition, there seems to be a problem with logical skeletons insofar as we can have judgments about those. As far as we can tell, logical skeletons correspond to second order formulas, with variables both in predicate and argument position. The second order formula $\forall X \forall x(Xx \rightarrow Xx)$ is true in all second order models and hence a logically valid second order sentence, as is the formula which results from removing the universal property quantifier, while $\exists X \exists y(Xy \wedge \neg Xy)$ is true in no second order models and so logically contradictory. Yet these are no more astonishing in terms of logic than first order logical validities. It would be strange if there were such a sharp division on logicality properties.

3 An alternative to L-triviality

Our idea is that the judgments about the examples in (1) are not the result of some property of logical absurdity or validity but rather of a failure of composition. While we agree with the proponents of L-triviality that the judgments about the uninterpretability of the examples in (1) stem from a problem with their semantics, we do not believe that L-triviality poses a viable solution. First, formulating precisely the notion of L-triviality faces serious obstacles as we have seen in the last section. Second, although the idea that truth and triviality have consequences for grammar is an interesting and provocative one, we believe that natural language examples do not provide evidence for it.

So for us, the judgments about the sentences in (1) have nothing to do with truth or triviality. Instead, we propose that the reason why examples in (1) are uninterpretable is that in building up the semantic representation for such examples, we encounter an insuperable semantic problem so that we can't fit the semantic pieces together to build up the full representation. This failure of composition comes about because some predicates in uninterpretable sentences presuppose that their arguments have a type or obey certain semantic principles that they in fact do not or cannot obey. This view thus claims that the sentences in (1) are uninterpretable for semantic reasons and provides an easy to understand intuition about their uninterpretability: we can't even say what the sentences in (1) mean or could mean because we can't put the meanings of their components together into a coherent whole.

3.1 Types for semantic composition

To flesh out this idea, we need to have a theory of the semantic principles that are operative in composition. These sort of questions have been pursued in formal semantic frameworks that use some notion of semantic typing to investigate cases of apparent meaning shift in phenomena like coercion and aspect selection (Cruse 1986, Jackendoff 1992, Nunberg 1993, Pustejovsky 1995, Asher 2011) *inter alia*. The idea of semantic types is familiar from Montague Grammar, but the authors just cited above attempt to extend the Church Montague system of functional types over the base types E (type of entities) and T (the type of truth values) to a much richer system incorporating subtypes of E and T, or PROP, the type of propositions.

But what sorts of meanings are types? TCL distinguishes two types of semantic content: *external* content and *internal content*. External content of an expression corresponds to the model-theoretic meaning that determines its appropriate extension (at points of evaluation). This is the usual notion of content in formal semantic theories. In addition, however, each word in TCL has a type. Types are semantic objects and encode the *internal meaning* of the expression associated with it. So for instance, the external semantics or extension of the word *wine* is a set of wine portions at some world and time, while the type or internal meaning of *wine* is given by the features we associate with wine — e.g., it's a liquid, a beverage, has alcohol and a particular taste). Internal semantics can also make use of multi-modal information; so olfactory and gustatory features can also play a role.

TCL's characterization of internal content yields a natural link between internal content and external, model-theoretic content. The internal semantics "tracks" the external semantics, in that in the majority of cases or in normal circumstances, the internal semantics determines appropriate truth conditions for sentences. The features encoded in the internal semantics enable speakers to correctly judge in normal circumstances whether an entity they experience falls under the extension of a term. The internal content given by the types doesn't determine the expression's extension in all cases, as philosophical, externalist arguments show (Kripke 1980; Putnam 1975). But assuming speaker competence, internal content should normally yield the correct extensions for expressions.[10],[11]

Before proceeding to introduce how composition works in this system, we briefly sketch what lexical semanticists and logicians assume about the structure of types. In a richly typed system, there are often difficulties if one assumes a universal type (Luo 1994), but it is standard to assume a most specific type, $\bot$, that is a subtype of all types above it. Thus, the set of types forms a semi-lattice ordered by the subtyping relation. For each syntactic category there is a maximal type. For instance, questions and propositions are both maximal types; there is no higher type that unifies the type of questions, which are sets or families of propositions, and the type of propositions. Similarly, there is no higher type above that of entity or E, though there are many subtypes of E. The same thing, we assume, holds of DP types and first order property types (Asher 2011). These highest types have semantic properties that are important for grammaticality. In the next section we see how DS buttresses this component of our view as well.

3.2 Composition and its failures

Types and internal content play an important role in our story about composition and failures of composition. Internal content is mainly responsible for guiding the construction of semantic logical forms. When composition succeeds, type presuppositions of predicate and argument are completely compatible, and a logical form for a sentence is constructed, all the internal meaning constraints given by the lexical expressions that make up that sentence have been satisfied given how those

[10]Note that internal content in TCL is not the same as the intension of an expression, if the latter is understood as a function from indices to truth values.

[11]Types, conceived as justifications, are formally modelled as defeasible proofs that can combine to create more complex types. This is possible by exploiting a deep relation between proofs and types known as the Curry-Howard correspondence (Howard 1980).

expressions compose together.

The process of integrating a lexical meaning into an interpretive context, can also shift a lexical meaning or introduce new meaning components as a result of the integration, depending on the type presuppositions of predicates and their arguments. Coercions such as (26) are an example of a construction where type presuppositions get satisfied, because the predicate licenses the addition of material needed to accommodate the type presuppositions of the argument; here the shift from *the artichokes* denoting a set of entities to it denoting events of eating artichokes.

(26) Marta enjoyed the artichokes.

But there are also cases of so called *semantic anomaly*, where composition fails because of an unreparable type mismatch. Some cases of semantic anomaly are well known and staple examples of introductory linguistic classes:

(27) #Colorless green ideas sleep furiously.

(28) #Tigers are Zermelo Frankel sets.

In such examples, it's relatively easy to see what's gone wrong; predicates like *are Zermelo Frankel sets* presuppose that their argument is some sort of abstract entity, while *tigers* contributes an argument whose semantic type must be a physical object. Given that these two types have incompatible individuation principles and essential properties (Asher 2011), the argument cannot compose with the predicate and we get a semantic anomaly. In these cases, the predicate does not license a coercion, and so the sentences like (27) and (28) are predicted to be semantically anomalous. And this gives us at least in principle a way of linking the ungrammaticality of (1) with semantic anomalies discussed in the coercion literature: type presuppositions can lead to irresolvable type conflicts when the type presuppositions of a predicate and its arguments are fundamentally incompatible. In that case the construction of logical form fails, the composition of meaning cannot go through, and we have a sentence that doesn't make sense.

Semantic anomaly is a graded phenomenon, see for example Magidor (2013). Some examples are interpretable in the right contexts but difficult outside of them like:

(29) Squirrels are human.

Some, like (27) or (28) are difficult to interpret in any context. Nevertheless, even with these examples, we can glean what could have been said. The contradictions between type presuppositions in these examples is somehow localized. And by relaxing those type presuppositions, for instance by supposing that the predicate in (28) simply is seeking an entity of general type E, the composition could actually have succeeded. Context might shift type presuppositions, but only to a limited degree.

The problems of composition in (1) are worse in that no relaxation of the type presuppositions seems possible. Not only are the semantic principles involved in the composition of the examples in (1) unshiftable in practice, we can't even comprehend how they could be shifted. They are somehow constitutive of the construction in a way that simpler type presuppositions of open class expressions, nouns, verbs and adjectives, are not. Our hypothesis is that the type presuppositions that are violated in (1) are constitutive of the type of denotation at the highest level of the type hierarchy. For instance, it is something about the semantic type of questions that leads to the uninterpretability of weak island sentences; it is something about the semantic type of constructions of the form *A but B* that leads to the uninterpretability of certain exceptive constructions; it is something about the nature of quantification that leads to the uninterpretability of *there is every girl*.

3.3 Meaning shift in co-composition

To get clearer on this intuition, we need to understand what it is to have a shiftable (internal) meaning, and if the hypothesis that shiftability depends upon the structure of the type system, what that structure is. In previous work (Asher et al. 2016), we have shown that there is shifting even when the type presuppositions of predicate and argument are completely compatible. While phenomena like coercion and aspect selection might seem like rather special linguistic phenomena, another sort of meaning shift is very common. For example, the adjective *heavy* has slightly different senses in each of the examples below:

(30) a. *heavy* box
b. *heavy* bleeding
c. *heavy* rain
d. *heavy* smoker

Pustejovsky (1995) calls such meaning shifting compositions *co-compositions*.[12] (Asher (2011) shows that the type shifting in co-composition is plausibly different from coercion.) There is no reason to suppose that type presuppositions, which are very general, are involved here. However, in addition to general type presuppositions, an expression in TCL also has a more specific, "fine-grained" type that encapsulates the internal content specific to the term. It is this fine-grained content that TCL exploits in co-composition. In TCL the noun and adjective meanings affect each other, and the output of an adjective-noun composition is the conjunction

[12]Note that the issue is not simply finding the right scale or a contextually specified cutoff on the scale for heaviness (as in *heavy mouse* vs. *heavy elephant*. Intuitively, what we need to capture is that the 'flavour' of the adjective changes depending on the context. Modeling this variation as pervasive ambiguity (e.g. via disjoint union types in a type system) is clearly unattractive as it would not capture that there is a 'common core' of the meaning of *heavy* that seems to be present in all cases. A disjoint union type might be right for homonymously ambiguous expressions (such as *bank*) but not for logically polysemous ones, expressions whose senses have some logical or metaphysical connection.

of a modified adjectival meaning and a modified noun meaning, which are both first order properties and apply to individuals, as in (31).[13] The *adjective-noun composition* schema (31) introduces functors that potentially modify both the adjective and the noun's internal content in co-composition and then conjoins the modified contents. In the schema below, A is the adjective, N the noun, $\mathscr{O}_A$ the functor on the noun given by the adjective and $\mathscr{M}_N$ the functor on the adjective induced by the noun:

(31) $\lambda x\, (\mathscr{O}_A(N)(x) \wedge \mathscr{M}_N(A)(x))$

Even logical truths have shiftable content. The predicates in them can undergo meaning shift that render informative an otherwise logical trivial sentence (cf. Chierchia and McConnell-Ginet 1992, Kamp and Partee 1995, Abrusán 2014).

(32) a. This table is red_1 and not red_2.
b. Every woman_1 is a woman_2.

In the above examples, we can (non-uniformly) shift certain aspects of the meaning of the sentence to get rid of the triviality and make the sentence informative. In particular, the two occurrences of the lexical predicates *red* and *woman* could get slightly different interpretations, i.e. their meaning could be shifted to express a particular aspect of the meaning of the predicates involved.[14]

4 What types of meaning can shift?

TCL alone does not supply detailed information about particular types, which is crucial to determining meaning shifts, nor does it tell us anything about the meaning functors in co-composition. Distributional Semantics, a computational approach to natural language semantics, can throw new light on meaning shifts in co-composition, as we have shown in Asher et al. (2016). This paper outlined a

[13]TCL's approach to adjective-noun co-composition is quite different from a standard Montagovian approach. In standard semantic treatments, an adjectival meaning is a functor taking a noun meaning as an argument and returning a noun phrase meaning; composition is a matter of applying the adjective meaning as a higher order property to the noun meaning.

[14]One might take this to be the grounds of another explanation of the data in (1) (see Abrusán 2014, Del Pinal 2017). While the examples in (32-a) are rescuable from triviality because of the possibility of meaning shift, in examples such as the ones in (1), triviality results from aspects of the meaning that cannot shift. Of course, this proposal assumes that some form of logical triviality is still the key to the ungrammaticality of the examples in (1). We don't believe that, and we think this explanation faces problems, as we could insist on making the two predicates in (32-a) be identical, as in:

(i) This table is red_1 and not red_2 and the property red_1 is identical to the property red_2.

Our criticism takes identity to be a logical operator, which already conflicts with Gajewski's proposal, but we believe that treating identity along with the truth functional and quantification expressions of first order logic on a par is defensible. If this is right then examples like (i) pose another problem for solutions that attempt to explain (1) through appealing to logical triviality.

close correspondence between TCL and DS methods. Further, results of our more recent work suggest that DS can also help us distinguish which aspects of meaning can shift and which ones cannot (cf. Abrusán et al. 2018).

The view from the DS approach connects to a growing body of work that assumes that the meaning of lexical words can be shifted or modulated in one way or another: either within the semantics (cf. e.g. Martí 2006, Stanley 2007, Asher 2011, Alxatib and Sauerland 2013) or within the pragmatics (Kamp and Partee 1995, Recanati 2010, Lasersohn 2012). Since we assume that meaning shift diagnosed by DS approaches happens at the (co-)compositional level, the view from DS is more in line with semantic approaches.

In this section we first briefly sketch how to compute meaning shifts more precisely using methods from computational semantics. Then we proceed to speculate about the question of shiftable vs. unshiftable meanings. We argue that DS can help us diagnose this as well. Aspects of the meaning that correspond to (or interact with) semantic dimensions uncovered by distributional semantics methods are in principle shiftable. In contrast, aspects of the meaning that are invisible for DS are unshiftable. We propose that semantic anomaly occurs only with unshiftable content. The reason why different types of semantic anomaly give rise to different intuitions is because unshiftable contents come in various flavors.

Distributional semantics can pick up the aspects of lexical meaning that vary with the context: these are the aspects of the meaning that are affected by changes in the distribution. They can perhaps also pick up aspects of possible variation at least among supertypes. However, those that can't possibly vary without destroying the highest type are (by definition) not sensitive to changes in the context (even counterfactual ones) and so are not discoverable by distributional methods.

In practice this means that there are (at least) two reasons why certain aspects of meaning are not shiftable: (a) The meaning is present in all contexts, and so it is invisible for DS; it will not show up in dimensions of the latent space where certain contexts are operative; (b) The meaning is attached to a type so high in the hierarchy that the type has no neighbours that share the same syntactic/semantic dependencies and so there is *nowhere in the space* for its meaning to shift. Such a meaning is also invisible to DS methods, but for a different reason. As a result, context invariant aspects come in different flavours.

Thus we argue that understanding the nature of meaning shifts helps us explain the intuitive difference between classic examples of semantic anomaly and the examples in (1). We close this section by applying our idea to the two examples discussed in Section 2: exceptives and weak islands.

4.1 Distributional semantics and meaning shifts

Let us provide a very quick introduction to DS and briefly describe a particular approach to capturing meaning shifts within this framework, Asher et al. (2016).

DS is based on the so-called "distributional hypothesis" by Harris (1954), according to which one can infer a meaning of a word by looking at its context. One way of thinking about word meaning within DS is to assume that it is a vector in some space **V** whose dimensions are contextual features. For example, the meaning of the word *raspberry* might be given by the vector that captures its co-occurence frequencies with all the words and/or grammatical features or dependency relations within a predefined context window in a corpus. Recording the vector meaning for each word (and possibly grammatical features) results in a word by context matrix. Such matrices are very large and very sparse. In order to bring out the 'information content' in them, dimensionality reduction techniques are applied. Dimensionality reduction reduces the abundance of overlapping contextual features to a limited number of meaningful, latent semantic dimensions. One such technique is non-negative matrix factorisation (NMF; Lee and Seung, 1999). As it turns out, reducing word-context matrices using NMF is particularly useful for finding topical, thematic information: the latent dimensions brought out by NMF can be interpreted as semantic features, or topics. Factorisation also allows a more abstract way of representing the meaning of a word: we can now say that the meaning of a word is represented by a vector of size k whose dimensions are latent features.

Thus DS can generate vectors to capture individual word meaning and bring out latent dimensions that might correspond to semantic features. But in order to capture meaning shift as in the examples in (30), the meaning of the adjective needs to be adapted to the context of the particular noun that it co-occurs with. In the TCL approach this means that the distributional model needs to provide us with the functors $\mathscr{O}_A$ and $\mathscr{M}_N$. In Asher et al. (2016) we have chosen two different approaches that meet this requirement: one based on *matrix* factorization (Van de Cruys et al., 2011) and one based on *tensor* factorization (Van de Cruys et al., 2013).

For example, the approach based on tensor factorization that we applied factorizes a three-way tensor[15] that contains the multi-way co-occurrences of nouns, adjectives, and other dependency relations (in a direct dependency relationship to the noun) that appear together at the same time. A number of tensor factorization algorithms exist; we opted for an algorithm called Tucker factorization in which a tensor is decomposed into a core tensor, multiplied by a matrix along each mode.

Given the results of this factorisation, we proceeded to compute a representation for a particular adjective-noun composition. In order to do so, we first extracted the vectors for the noun ($\mathbf{a}^i$) and adjective ($\mathbf{b}^j$) from the corresponding matrices **A** and **B**. We multiply those vectors into the core tensor, in order to get a vector **h** representing the importance of latent dimensions given the composition of noun i and adjective j. By multiplying the vector representing the latent dimension with the transpose of the matrix for the mode with dependency relations ($\mathbf{C}^T$), we are able to compute a vector **d** representing the importance of each dependency feature given the adjective-noun composition. The vector **d** is in effect the DS version of TCL's functor $\mathscr{O}_A$, which we now have to combine with the original noun meaning. This last step goes as follows in DS: we weight the original noun vector

[15] A tensor is the generalization of a matrix to more than two axes or *modes*.

according to the importance of each dependency feature given the adjective-noun composition, by taking the point-wise multiplication of vector **d** and the original noun vector **v**.

Finally, observe an example illustrating the unshifted meaning of the adjective *heavy* vs. the shifted meaning of the same adjective in the context of the noun *traffic* as computed by our tensor method. In the examples below we list the ten closest adjectives (as computed by cosine similarity) to the unmodified and the modified adjective, respectively:

(33) $\mathbf{heavy}_A$: $heavy_A$ (1.000), $torrential_A$ (.149), $light_A$ (.140), $thick_A$ (.127), $massive_A$ (.118), $excessive_A$ (.115), $soft_A$ (.107), $large_A$ (.107), $huge_A$ (.104), big_A (.103)

(34) $\mathbf{heavy}_A$, traffic_N: $heavy_A$ (.293), $motorised_A$ (.231), $vehicular_A$ (.229), $peak_A$ (.181), $one\text{-}way_A$ (.181), $horse\text{-}drawn_A$ (.175), $fast\text{-}moving_A$ (.164), $articulated_A$ (.158), $calming_A$ (.156), $horrendous_A$ (.146)

There is an evident shift in the composed meaning of *heavy* relative to its original meaning; there is no overlap in the lists (33) and (34) above except for *heavy*. We see this also in the quantitative measure of cosine similarity, sim_{cos} between the original vector for *heavy* $\vec{v}_0$ and the modified vector for *heavy* $\vec{v}_1$ as modified by its predicational context: With the tensor model, on average, $sim_{cos}(\vec{v}_{orig}, \vec{v}_{mod})$ was 0.2 for adjectives and 0.5 for nouns.

4.2 Constraints on shiftability from DS

The distributional method that we described above for calculating meaning shift adapts the vector of the original predicate to its predicational context using the latent dimensions derived during dimensionality reduction. All of this depends on the fact that the original predicate's distributional meaning, the functor on it, and the result of the application of the functor to the distributional meaning, are defined in the same vector space. This has immediate consequences for types that are at a maximally general level, like questions, whose type is a family or set of propositions, propositions themselves, the general type E of entities, the general type of determiner phrases DP, quantifiers or second order properties, the general type of first order properties and so on. These elements don't have any neighbours in a vector space, in the way that say a common noun like *traffic* does, because other expressions that are not of say DP type will belong to a different syntactic category with different syntactic/semantic dependencies; they will be in a different space (though they may share the same latent dimensions as particular DPs). As a result, there is *nowhere in the space* for such a type to shift and preserve its corresponding syntactic category. DS methods *can't shift* those types, in the way that it can shift *heavy* in the context of *traffic*. So these types perforce have semantic principles that are invariant, if they have any semantic principles at all. Furthermore, because these types don't have neighbours in the vector space if indeed they inhabit a vector space at all, these semantic principles will not show up in particular latent dimensions of the vector space. DS methods, at least of the sort we have employed, will not be able to see these principles. For expressions that have type presuppositions *at this level*, it follows that they cannot be shiftable.

Of course not all type presuppositions are of such a general type; many as we have seen are subtypes of E. Nevertheless, DS puts constraints on the shiftability of these types as well. From our studies of adjective-noun composition we have found that almost all modification of noun meanings is subsective; that is, either an ADJ N is an N or it looks like an N, smells like an N, etc. We hypothesize then that non-subsective type shifts are not allowed in co-composition. Given that type presuppositions are very general, requiring a shift over type presuppositions would require in effect a non-subsective type shift — e.g. from a physical object to an abstract object type. We note that this is not necessarily the case for adjectival type shifts, only for type shifts involving nominals and the basic types of objects. If this is not allowed, then we would predict that most normal type presuppositions cannot be shifted during composition. That is, one cannot shift type presuppositions to get composition to succeed (we've already seen that coercion in TCL isn't modelled as a shifting of types of the argument or predicate, but rather a shift in the predicational environment).

On the other hand, one can imagine another kind of shift — one that rescues some sort of content from a predication in which there is a type clash between a predicate and its argument. (27) or (28) are examples of this. If we simply move to the supertype of the type presuppositions, we can see that the author of (27) or (28) was predicating a property of some object; she was just confused about or wilfully misusing the meaning of the property or the object expression. The distance between the type presuppositions of the predicate and arguments may be great, but it is still defined, as *tiger* and say *ordinal* are both in the same vector space DS defines for common nouns. And thus DS and TCL together provide us with a means to distinguish (27) and (28) from the examples in (1).

4.3 Logical meaning and DS

Logical meaning is present in all contexts, and so we expect it to be invisible for DS; in particular we expect that it will not show up in dimensions of the latent space where certain contexts are operative. The way the distributional method calculates meaning shifts implies that meaning shift crucially depends on the latent dimensions that we find during tensor factorisation: it is the semantic features implicitly present in the latent dimension that drive the meaning shift. Thus whether or not we get logical meaning to shift depends on whether we find latent dimensions with our dimensionality reduction methods that correspond to logical meaning. In recent work (cf. Abrusán et al. 2018) we performed a number of preliminary experiments similar to the ones described in Section 4.1 but this time with determiner-noun compositions. Specifically, we looked at four determiners, *a, any, some* and *every* using two different corpora: Wikipedia, and a corpus of unpublished novels collected from the web (Zhu et al., 2015).[16] In the resulting factorization, determiners and nouns as well as dependency relations were all linked to same latent dimension. An intuitive evaluation of the semantic coherence of each of the 30 dimensions was conducted, and we have found that many of these seem to capture interesting semantic features, albeit not logical features.

For example, in the case of *any* we get a dimension that captures its peculiar distribution. Most interestingly, perhaps, the dimensions we find with the quantifier *some* correspond to non-logical aspects of its use that have puzzled semanticists since a long while. The first of these is uncertainty about identity, also known as the *epistemic* aspect of indefinites, cf. Kratzer and Shimoyama (2002), Alonso-Ovalle and Menéndez-Benito (2015), among others. Other aspects of the determiner *some* include measure and kind readings. In contrast, in the case of the determiner *a*, we mostly found topical dimensions, e.g. legal, publishing, building construction, political campaigns, people, etc; in some dimensions *a* appeared within prepositional modifier phrases (*in a chair, with a grin*) and the rest of the dimensions were uninterpretable to us.

What we have described above is still work in progress, but it is already clear that we are not getting any dimensions via tensor factorisation that correspond to logical meaning. As a consequence, we are not going to get logical meaning to shift. This is not surprising given that logical meanings are supposed to validate logical deductions universally regardless of context. Thus the fact that logical meaning shouldn't shift with content comes with the definition of logical meaning and the universally valid inferences it purports to underwrite. On the other hand the dimensions that we do get correspond to the lexical/distributional aspects of the conceptual meaning of quantifiers. In the light of this, one way to interpret our results with the determiner *a* is that this determiner does not have any extra conceptual content beyond its logical meaning.

These results suggest that we cannot distinguish logical and non-logical items based on invariance with respect to meaning shift: both lexical and functional words can have shiftable (i.e. co-composing, context sensitive) aspects to their meaning and also stable (not-co-composing, context-invariant) aspects.

4.4 Examples

We are now in the position to come back to the examples discussed in Section 2 and examine them in the light of our proposal.

Exceptives As we have seen in Section 2, *but*-exceptives are unacceptable with certain quantifiers:

(35) *Some boy but John smokes

[16] The former corpus contains about 1 billion words, the latter about 1,5 billion words.

Using TCL's internal contents and notion of a type constrained composition gives a different sort of way to use unshiftable contents to predict the ungrammaticality or rather uninterpretability of the sentences in (1). Consider this alternative explanation to that of von Fintel (1993) in the case of exceptive constructions. For simplicity, we'll look at the special case of exceptive constructions where *but* links two DPs, as in DP_1 *but* DP_2, where the second is a proper name or definite description. But we can generalize the following analysis to the more general pattern. Notice that what von Fintel says about the construction amounts to the following entailment:

(36) $[everyA\ but\ B](C)$ holds iff $every(A \setminus B, C)$ and $every(A \cap B, \neg C)$.

We also have:

(37) $[NoA\ but\ B](C)$ holds iff $No(A \setminus B, C)$ and $No(A \cap B, \neg C)$.

As a general principle then we should have:

(38) $[Det_1 A\ but\ B](C)$ holds iff $Det_1(A \setminus B, C)$ and $Det_1(A \cap B, \neg C)$.

That is, the exceptive construction entails that a property C is the argument of DP_1 *but* DP_2 iff $\{A \setminus B, A \cap B\}$ forms a partition of A that corresponds exactly to the partition of A, $\{C \cap A, C^c \cap A\}$, where C^c is the complement of C. But this means that the determiner in DP_1 must distribute to form a partition over A when it combines with DP_2; i.e.,

(39) for every C, if $Det_1(A \setminus B, C)$ and $Det_1(A \cap B, \neg C)$, then every element of A is determined with respect to C.

We can say in this spirit that DP_1 *but* DP_2 holds of properties that induce a partition over the restrictor set A, what we could call Ramsey properties. This puts a type presupposition on the determiner that A combines with, as well as a type presupposition on A and B. In particular, it presupposes the type of object that is a B must be a subtype of the type of object that is an A and that whatever type of determiner A combines with must entail a partition of A in order for composition to succeed, for the Ramsey property to be built. But only determiners that are downward entailing in the restrictor like *every* and *no* can do this. Any other determiner will not set up the partition match and so *A but B* will not compose with them. According to this explanation, the problem with exceptive constructions is a type conflict that cannot be resolved ('shifted away') during composition: the type of the Ramsey property is constituted by the invariant meaning of determiners and their monotonicity properties. If the conflict cannot be resolved, then composition fails and so we predict the sentence to be uninterpretable.

We note that this type conflict is of a much more abstract nature than the one underlying (27) or (28). For those type conflicts, we can still imagine shifting functors like those in co-composition that would shift the noun *tiger* say to some radically different type of object, like *number* or *ordered pair*. The type clash here in TCL is immediate, because the type of the argument, in this case the individual variable introduced by the subject's head noun, is transmitted to the predicate in composition; and the incompatibility is simply checked. It's also relatively obvious how to fix it, even though type shifting can't shift the type outside of the space of types covered by the noun without generating some at least mild form of semantic anomaly: by allowing the co-composition functors to do their work on the head noun of the subject. Our functors make sense at an individual word level, and this is compatible with the idea that type checking for semantic well-formedness is a *local phenomenon*. With the exceptive construction, the type clash has to do with a much more abstract property of determiners, monotonicity, and further the type of the determiner is not transmitted to the conjunction; only the type of DP is. Given the nature of type presuppositions, we then *can't* check composition locally. And in fact, *some boy* could compose with *but* (as in *some boy but no girl*); but given that it has a positive DP in its second argument, the composition will fail because we have an exceptive construction and the type of the determiner can't yield the right input to *but*. Without this local checking, we can't invoke the shifting functors to try to make sense of the exceptive construction. And this explains why such sentences are so mysterious, even in comparison with semantically anomalous sentences like (27) and (28).

Weak Islands As for weak islands, let's consider the case of negative weak islands as an illustration. We can almost use Abrusan's account verbatim, appealing to type conflicts instead of logical contradictions. Recall that (21) has a true answer that entails all the true answers if and only if John's height is 0, which in turn means that (21) can be felicitous only when it is common knowledge that John's height is 0. This already conflicts with the requirement on the question type at issue, since such a question has in its answer set meaning only non 0 answers. Hence, composition fails and we cannot construct a coherent logical form for sentences like (21).

The only thing we need to check in this explanation is that the derivation of the presupposition for (21) indeed follows from internal meaning postulates. In particular we can impose constraints on questions types *à la* Dayal: a question type, which is a set of propositional types, must have a uniquely most informative element. As deduction is reflected in internal meaning via proof theoretic rules, if we interpret this constraint on questions proof theoretically, we can derive the presupposition for (21) in terms of internal unshiftable meanings. That presupposition understood as a type restriction imposes constraints on the types of other variables and conflicts with the type requirements of questions. Once again as in the case of exceptives, the type clash can't be diagnosed locally. In consequence, we can't make use of our shifting functors to rescue the sentence or counterfactually reconstruct a good semantic composition.

The explanation for the semantic incoherence of positive weak islands is a bit different but follows the general outlines of the type conflict sketched for negative weak islands.

5 Conclusion

In this paper, we've taken a new look at certain types of unacceptable sentences. We've argued that these sentences are ungrammatical for semantic reasons, but our analysis differs from that of Gajewski (2002) and those following him in that we locate the semantic problem at the level of semantic composition, not at the level of logicality. We have argued that our position is preferable for several reasons, and we have illustrated this approach with a brief analysis of weak island sentences and exceptive constructions. What we find really interesting about our approach, however, is what it tells us about meaning and composition. Unacceptable sentences on our story offer evidence for a rich system of types that semantic composition makes use of. And they force us to investigate why, when meanings apparently can and do shift with context, some meanings can't. In this investigation, we have fashioned an interesting and fruitful marriage of symbolic and statistical techniques for analyzing meaning shift. The picture of meaning that emerges cuts across old distinctions like the lexical/ functional content distinction and borrows both from statistical and formal notions.

Acknowledgments

We are grateful to the editors of this volume for the invitation and to the audience at Sinn und Bedeutung 22 (Berlin/Potsdam, September 2017) for helpful comments. The research reported here was supported by a Marie Curie FP7 Career Integration Grant, Grant Agreement Number PCIG13-GA-2013-618550, a grant overseen by the French National Research Agency ANR (ANR-14-CE24-0014), and by ERC grant number 269427.

References

Abney, S. (1987). *The English Noun Phrase in Its Sentential Aspect.* Ph. D. thesis, MIT.

Abrusán, M. (2007). *Contradiction and Grammar: the Case of Weak Islands*. Ph. D. thesis, M.I.T.

Abrusán, M. (2014). *Weak Island Semantics*. OUP Oxford.

Abrusán, M., N. Asher, and T. V. de Cruys (2018). Content vs. function words: The view from distributional semantics. In *Proceedings of Sinn und Bedeutung 22*.

Abrusán, M. and B. Spector (2010). An Interval Based Semantics for Negative Degree Questions. *Journal of Semantics*. doi:10.1093/jos/ffq013.

Alonso-Ovalle, L. and P. Menéndez-Benito (2015). *Epistemic indefinites: Exploring modality beyond the verbal domain*. Oxford University Press, USA.

Alxatib, S., P. P. and U. Sauerland (2013). Acceptable contradictions: Pragmatics or semantics? *Journal of Philosophical Logic 42*, 619–634.

Asher, N. (2011). *Lexical Meaning in Context: A web of words*. Cambridge University Press.

Asher, N., T. van de Cruys, A. Bride, and M. Abrusán (2016). Integrating type theory and distributional semantics: a case study on adjective-noun compositions. *Computational Linguistics 42*(4), 703–725.

Barwise, J. and R. Cooper (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy 4*(2), 159–219.

van Benthem, J. (1989). Logical constants across varying types. *Notre Dame Journal of Formal Logic 30*(3), 315–342.

van Benthem, J. (2002). Invariance and definability: two faces of logical constants. essays in honor of sol feferman. In W. Sieg, R. Sommer, and C. Talcott (Eds.), *Reflections on the Foundations of Mathematics.*, ASL Lecture Notes in Logic 15., pp. 426–446.

Bonnay, D. (2006). *Qu'est-ce qu'une constante logique?* Ph. D. thesis, University Paris 1.

Bonnay, D. (2008). Logicality and invariance. *Bulletin of Symbolic Logic 14*(1), 29–68.

Bonnay, D. and D. Westerstahl (2012). Consequence mining. *Journal of Philosophical Logic*, 1–39.

Chierchia, G. (2013). *Logic in grammar: Polarity, free choice, and intervention*. OUP Oxford.

Chierchia, G. and S. McConnell-Ginet (1992). *Meaning and Grammar: An Introduction to Semantics*. MIT Press.

Cruse, D. A. (1986). *Lexical Semantics*. Cambridge University Press.

Dayal, V. (1996). *Locality in WH quantification: Questions and Relative Clauses in Hindi*. Studies in Linguistics and Philosophy. Kluwer Academic Publishers.

Del Pinal, G. (2017). The logicality of language: Triviality and logical form. *Noûs (forthcoming)*.

Feferman, S. (1999). Logic, logics, and logicism. *Notre Dame Journal of Formal Logic 40*(1), 31–54.

von Fintel, K. (1993). Exceptive constructions. *Natural Language Semantics 1*(2), 123–148.

von Fintel, K. (1995). The formal semantics of grammaticalization. In *Proceedings of NELS 25*, pp. 175–189.

Fox, D. and M. Hackl (2007). The Universal Density of Measurement. *Linguistics and Philosophy 29*, 537–586.

Gajewski, J. (2002). On Analyticity in Natural Language. Ms., MIT.

Gajewski, J. (2008). More on quantifiers in comparative clauses. In *Semantics and Linguistic Theory*, Volume 18, pp. 340–357.

Gajewski, J. (2009). L-triviality and grammar. Handout, UConn Logic Group, University of Connecticut.

Heim, I. and A. Kratzer (1998). *Semantics in Generative Grammar*. Oxford: Blackwell.

Howard, W. A. (1980). The formulas-as-types notion of construction. In P. Seldin and J. R. Hindley (Eds.), *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, pp. 479–490. Academic Press. First appeared in **Language**, vol. 29 (1953), pp.47-58.

Jackendoff, R. (1992). *Semantic structures*. MIT press.

Kamp, H. and B. Partee (1995). Prototype theory and compositionality. *Cognition 57*(2), 129–191.

Kratzer, A. and J. Shimoyama (2002). Indeterminate pronouns: The view from Japanese. *The Proceedings of the Third Tokyo Conference on Psycholinguistics*, 1–25.

Kripke, S. A. (1980). *Naming and Necessity*. Cambridge, MA.: Harvard University Press.

Kuno, S. and K. Takami (1997). Remarks on negative islands. *Linguistic inquiry 28*(4), 553–576.

Lasersohn, P. (2012). Contextualism and compositionality. *Linguistics and Philosophy 35*, 171–189.

Lee, D. D. and H. S. Seung (1999). Learning the parts of objects by non-negative matrix factorization. *Nature 401*(6755), 788–791.

Luo, Z. (1994). *Computation and Reasoning: a Type Theory for Computer Science*. Oxford, U.K.: Oxford University Press.

Magidor, O. (2013). *Category Mistakes*. Oxford University Press.

Martí, L. (2006). Unarticulated constituents revisited. *Linguistics and Philosophy 29*.

Nunberg, G. (1993). Indexicality and deixis. *Linguistics and Philosophy 16*(1), 1–43.

Pustejovsky, J. (1995). *The Generative Lexicon*. MIT Press.

Putnam, H. (1975). The meaning of 'meaning'. In K. Gunderson (Ed.), *Language, Mind and Knowledge. Minnesota Studies in the Philosophy of Science, vol. 7*, Minneapolis, MN, pp. 131–193. University of Minnesota Press.

Recanati, F. (2010). *Truth-conditional pragmatics*. Clarendon Press Oxford.

Rizzi, L. (1990). *Relativized Minimality*. MIT Press.

Schwarzschild, R. and K. Wilkinson (2002). Quantifiers in comparatives: A semantics of degree based on intervals. *Natural Language Semantics 10*(1), 1–41.

Sher, G. (1991). *The Bounds of Logic*. Cambridge, Mass: MIT Press.

Stanley, J. (2007). *Language in context: Selected essays.* Oxford: Clarendon Press.

Szabolcsi, A. and F. Zwarts (1993). Weak islands and an algebraic semantics for scope taking. *Natural Language Semantics 1*, 235–284.

Tarski, A. (1966/1986). What are logical notions? *History and Philosophy of Logic 7*(2), 143–154.

Van de Cruys, T., T. Poibeau, and A. Korhonen (2011, July). Latent vector weighting for word meaning in context. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, Edinburgh, Scotland, UK., pp. 1012–1022.

Van de Cruys, T., T. Poibeau, and A. Korhonen (2013). A tensor-based factorization model of semantic compositionality. In *Conference of the North American Chapter of the Association of Computational Linguistics (HTL-NAACL)*, pp. 1142–1151.

Zhu, Y., R. Kiros, R. S. Zemel, R. Salakhutdinov, R. Urtasun, A. Torralba, and S. Fidler (2015). Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. *CoRR abs/1506.06724*.