**Notes on Reviewer Commentary**

Since a number of the comments were of a more general nature, I'll start with a brief overview of the major changes, and how this addressed their comments.

The paper has a new organization as follows:

1. Introduction
2. Canonical Typology
   1. Why Canonical Typology?
   2. Key concepts in Canonical Typology
   3. Prototypes versus canonical points, and some clarifications
3. Categories and the lexeme
   1. Crosslinguistic and language-specific categories
   2. What's being categorized: Defining the lexeme
   3. Approaches to categorization
   4. Rigid and flexible categories
4. Flexible categories: The range of the phenomenon
5. Conclusion

The introduction now spends more time discussing different approaches to categorization, and explaining where the present article fits in among these approaches. It motivates the central problem - the deadlock that arises from debates on lexical flexibility - and discusses how Canonical Typology is well-suited to address it. In general, I now take the time throughout the article to situate this paper and its ideas in the extant literature on the topic, showing how it relates to and is different from other approaches (thus addressing some of Reviewer 3's concerns in that regard.)

Section 2 spends more time motivating Canonical Typology, and clearing up some common points of confusion, including a number pointed out by reviewers. I note cases where the canonical approach has been used in the literature, and briefly exemplify a few of them. I devote one section to explaining CT's basic concepts and methodology, and another to discussing its similarities/differences with prototype theory. The result is that this paper now also serves as a helpful introduction to anybody unfamiliar with the method.

Reviewer 2's comments regarding the lexeme motivated a number of changes. The reviewer was quite right that the entire framework of the paper hinges crucially on one's concept of the lexeme, so I now give an explicit definition of the lexeme in Section 3, discuss alternative definitions, and present a brief canonical typology of the lexeme, which will motivate many of the criteria later in the paper. The section at the end of the previous version, on the semantics of lexical flexibility, has now been subsumed partly under this section, and partly under Criteria 6 and 7, which deal explicitly with semantics. This prevents the discussion of semantics from being broken up, and adds a good deal more coherence to the paper. Section 3 now also includes a much more explicit presentation of Hengeveld's parts-of-speech typology, examines its theoretical foundations, provides a critique, and then establishes an alternate conception of flexible categories for use in this paper. It also takes the time to consider previous critiques of this approach (especially Croft's), and shows how my revised notion of lexical flexibility addresses these problems. All this again helps to address Reviewer 3's comments. It should be much clearer now how the present approach is an improvement on previous ones, and how it fits into the literature.

Section 4, where I lay out the ten criteria, is largely the same in structure, although much more detailed in its treatment. In particular, I've now included numerous examples illustrating how languages might be flexible or rigid according to each criteria, giving the reader an idea of the range of the possible phenomena. (More examples are now used in the other sections as well.) I also specifically note where my criteria match those of other writers on this topic, and in a number of cases have adjusted my terminology to be more precise. In fact, this is a general change - making sure explicitly define any terms used, and use them consistently throughout. The criterion on semantic shift, one of the most important issues in the debate, has been split into two separate criteria, and the discussion is more extensive, and deals explicitly with the issue of idiosyncratic semantic shift, showing how this is not a problem for the canonical typology presented here (utilizing the definition of the lexeme presented earlier). Finally, two criterion have been switched so as to be ordered more logically, and the prior Criterion 10 has been removed (still leaving me with 10 criteria, since Criterion 6 was split in two).

I conclude with a brief summary of the criteria that readers should find very helpful, and then show how these ten criteria can be successfully applied to tackling the motivating problem of this paper - ending the deadlock in the debates on lexical flexibility. I apply the typology presented here to the Mundari debate, and assess Mundari for each one of the criteria. I show that Mundari is flexible by some criteria, rigid by others, and predominantly flexible overall. The exercise illustrates the utility of assessing lexical flexibility along many different criteria, rather than simply providing an affirmative or negative answer to the question, 'Is Mundari a flexible language?'. The brief demonstration tries to illustrate how researchers might usefully apply the criteria laid out here, and perhaps even quantify lexical flexibility for different languages.

One final note: I have an email out to Dirk Geeraerts asking for permission to use the illustration in Figure 5, but have not heard a response yet. If I don't receive a positive response from him, I can replace this with a diagram of my own.

Specific reviewer comments are below, with my responses. I've changed the locations so that they reference locations in the revised document instead of the original.

**Reviewer 1**

This paper does a very interesting and timely job in drawing together several strands of research on (the problems associated with) lexical categorization and cross-linguistic comparison of lexical categories. The structure and style of the paper are clear. However, in several places the argumentation is too implicit and the connections with earlier research are too loosely drawn. This makes it difficult to evaluate what the merits and the innovative value are of the proposed Canonical approach, in relation to other functional-typological approaches, and what the broader theoretical and empirical implications are of adopting the Canonical approach. Hopefully the comments below are helpful in amending this – more examples would be good, too.

Connections to other research now much more explicit throughout. CT is compared and contrasted to similar approaches in the Introduction and Section 2. The paper now contains a multitude of examples (55 in all, I think).

There is some terminological inconsistency here: ‘word class’ is used instead of ‘lexeme class’, in order to encompass both lexical and functional categories, but then ‘lexeme’ is used instead of ‘word’ in order to distinguish a lexical entry (=’root’) from its morphologically marked instantiations. Perhaps ‘lexical class’ is better than ‘word class’, as functional categories do not seem to be at issue.

Terms are made explicit in Section 3.2 when defining the lexeme, and applied consistently throughout. It is also made clear that the focus of the paper is on lexical categories only, not word classes generally.

There is an important difference between Hengeveld and Rijkhoff’s approach on the one hand, and Croft’s/Haspelmath’s approach on the other hand: While H&R take into account only *pragmatic/syntactic functions*, Haspelmath and Croft rather use a combination of such functions with semantic meanings. Thus, for Hengeveld, an adjective is a lexeme that functions without further measures as a *modifier*in a referential phrase, but for Haspelmath and Croft an adjective also *denotes a property*. Especially for Croft’s theory of parts-of-speech, the cross-classification of pragmatic function and semantic meaning is crucial. I think this deserves some more attention here, including a justification of why the author does not take into account semantics as a comparative concept, or, for that matter any of the “numerous other comparative concepts one might bring to bear for the purpose of comparing word classes cross-linguistically”.

This is noted in Section 3.4 now, where I examine Rijkhoff's approach and critique it precisely for these differences, then revise the concept to be more in line with Croft/Haspelmath's approach.

Section 2.1 - I think the claims [regarding the capabilities of CT] are quite strong, but at the same time too vague: what kinds of “phenomena” have been “obscured and ignored” or unjustly treated as “identical”? These claims should either be more clearly illustrated and referenced, or they should be omitted.

Section 2.1 now quotes Corbett directly, stating the perceived gap in typological methods he was trying to address when he formulated CT, and explains how his work has met these goals.

 Section 2.3 - It remains difficult to imagine what causes a set of properties to be “expected”. What is the basis of this expectation, if it is not frequency or prototypicality? It would seem that a potential advantage of Canonical Typology over ‘non-theoretical typology’ is that the former makes predictions before looking at cross-linguistic data, rather than after doing so. But then it should be clearer where these predictions come from.

"Expected" was a poor choice of words, since CT isn't in the business of prediction. I make this important clarification in Section 2.3 now.

Last large paragraph before Section 4 - It is expected, on the basis of principles of functional transparency, that no word class will be canonically flexible. This seems to indicate that the canonical situation is at the same time the expected and the not expected situation. Also, the prediction that canonical lexical flexibility is ‘dis-functional’, i.e. requires some sort of counterbalance in order to preserve transparency is not new – some more references seem to be in order here (e.g. Hengeveld, Rijkhoff & Siewierska 2004; Rijkhoff & Van Lier, forthcoming; Frajzyngier & Shay 2002; Lehmann 2008).

The first comment should no longer be relevant, now that I've dispensed with the term "expected" and clarified the canonical approach more.

The second comment is indeed an important connection to make. I've added these references and spend a few sentences discussing this point.

A “comparative concept” is defined here as an intersection of specific properties or criteria. However, while comparative concepts are usually thought of as functional in nature (i.e. relating to either pragmatic or semantic substance), it seems that properties/criteria can be of a formal nature as well (such as the ability to take certain morphological marking). Could the status of “criteria” in Canonical Typology be made more explicit?

I've addressed this clarification explicitly with a paragraph in Section 2.3 '...some clarifications' now.

Section 3.3, 2nd paragraph - (“I do not take the distributional properties of a lexeme to be its essential characteristics”): What then, if not distributional properties (i.e. morpho-syntactic behavior relative to semantic/pragmatic functions) are the properties of a lexeme used for categorization? Notably, this is not to disagree with the methodological point that we should not privilege one method (or property) over another, but rather aims at clarifying what is meant by “distributional properties” and why they would not count for categorization.

My thinking here was that distributional properties might be necessary but not sufficient for categorization. That is, there might be non-distributional properties which are relevant to categorization as well, and so the distributional ones shouldn't be given priority. But after some more thought and rereading of Croft, I think Croft's conception of the distribution of a lexeme is broad enough to preclude the possibility of non-distributional categorization, and I have trouble imagining what a relevant non-distributional property would be. I note this possibility briefly but otherwise I've retracted this statement and I just follow Croft's distributional method now.

In Croft’s approach, lexical classes are defined by constructions: depending on which construction one looks at, the conceptual space of the lexicon is divided up in different ways. This is not the same, of course, as saying that a single lexeme can belong to different classes, but it does differ from the “lumping” approach, in that it does not “choose” between distributional criteria or “methods of classifying lexemes”. The Hengeveldian/lumping approach does choose: it takes into account only distribution over syntactic slots. Perhaps this could be made even more explicit?

This is addressed in Section 3.4 now, where I dissect Hengeveld's parts-of-speech typology.

Criterion 1, example 18 - (“still surprisingly flexible”): this seems an odd characterization, given that Criterion 1 states that the presence of derivational morphology is *not*characteristic of flexibility.

While I've abandoned the phrasing "surprisingly flexible", I use this opportunity to clarify that Inupiaq is extremely rigid according to the present criterion, but still flexible in that it can change propositional act functions at all. Consider the opposite case: a lexeme that isn't allowed to change functions even *with* derivational morphology. Clearly this would be a highly rigid lexeme. But since Inupiaq is not like this, it still exhibits some amount of flexibility.

Criterion 2 - There is a difference between (i) inflectional categories that are always expressed in combination with a specific syntactic function; no matter what the category of the lexeme filling the slot, and (ii) differences in the range of inflectional morphology that certain lexemes can express when used in a particular syntactic function. Is it the case that any word used for predication in Swahili takes TAM affixes and agreement, and any word used for reference takes noun class prefixes (corresponding to (i))? If so, the terms “noun” and “verb” in the sentence just above example (9) refer to syntactic classes, not lexical classes. Or do only lexical verbs in Swahili take TAM and agreement in predicative function (and nouns in predicative function show different behavior), while only lexical nouns take class prefixes in referential function (and verbs do not)? It is also worth pointing out, perhaps, that differences in inflectional behavior do not play a role in the Hengeveldian approach to categorization; only derivational morphology does. Finally, it seems that Criterion 2 does not take into account the possibility of expressing the same functional categories (for instance TAM) by means of particles, rather than inflection.

This is an excellent point. Put differently, is inflection a property of the syntactic slot, or of the lexeme? In fact, this is a great alternate way to phrase this criterion, so I do so in the article. I discuss this comment explicitly under criterion 2 now.

Criterion 5 “principle of exhaustiveness”: Evans and Osada also propose a criterion of exhaustiveness (see p.20), which is not the same as Criterion 5. Perhaps it is worth making this explicit at this point?

This has been noted in Footnote 13 now.

Assuming some process of accretion in meaning does not presuppose vague/underspecified lexical meaning; constructional meaning can also be added to a fully specified lexical meaning.

This comment was just due to a syntactic ambiguity on my part. It should have been read 'posit some accretion in meaning, OR say that the lexemes are underspecified or vague, etc. etc.' I've reworded this.

(Note 11): As mentioned later on, the issue of semantic enrichment is in fact hotly debated. The formulation of this note seems too weak, especially since in section 6 the debate is more clearly outlined and the author takes a more explicit position in relation to this debate. Perhaps, therefore, this note is unnecessary?

Agreed. Note removed. This is addressed more explicitly in Section 3.2 now anyway.

Criterion 7 (“We can explain…..agent.”): this type of explanation has only limited potential: the relevant difference between hammer and spy seems to reside in animacy. However, two inanimate items, such as hammer and rope may have different interpretations when used as verbs, which are not readily predictable from “the nature of the entities involved”.

There is now a long discussion on the issue of idiosyncrasy in semantic shift underneath Criterion 7, which explicitly acknowledges this reviewer's point - even exemplifying it with a few cases - and then explaining how the present canonical typology deals with this problem.

Section 4.2.2 (“across lexemes”): Does flexibility across lexemes not fall under Criterion 4?

"Across lexemes" refers to the number of lexemes in the lexicon which are considered flexible, whereas Criterion 4 states that flexible lexemes indistinguishable in the direction of their flexibility. I've tried to clarify this in both places.

p.18, example (18): This example what it necessary to be more explicit as to what is understood as a ‘function’ (see also my 2nd comment above): Here, the ability to fulfill 4 semantic functions is characterized as an extremely limited range of flexibility. This is probably because all of these functions fall inside the pragmatic/syntactic function of reference. On the other hand, if a lexeme is able to fulfill 3 different types of such pragmatic /syntactic functions, then it would be maximally flexible. Thus, the degree of flexibility is dependent on the type of function one is interested in, i.e. on the type of comparative concept that is taken into account.

This comment has been addressed mostly by being more clear about comparative concepts in my introductory sections and defining the lexeme more explicitly. Throughout the paper I've adopted Croft's propositional act functions in particular as my comparative concepts.

(Example 54: Cayuga) See Mithun’s (2000) counter-analysis, which crucially appeals to the effects of using multiple-level criteria for lexical categorization.

Mithun's analysis makes it clear that the omnipredicative status of Cayuga is purely apocryphal. For the purpose of clearing up this myth, I've kept the example and noted Mithun's critique, and then provided a different example instead.

Criterion 6 - It would be good to add references to (24c) and (24d) as well. From the text below (24), it seems that Spencer (2005) represents (24c). What is missing here too, is some reference to work (in the generative tradition, e.g. Arad 2003, 2005) on the distinction between root-based zero-derivation (corresponding to (25b)?) and zero-derivation from an already categorized lexeme (corresponding to (24c)?).

The reviewer is referring to the list of items I-IV on page 44, which has now been restructured. There is a great deal of terminological confusion in the literature here, again circulating around the notion of a lexeme, so I've spent a paragraph discussing the different stances scholars adopt in regards to the semantics of flexible categories, and different ways the terms polysemy and heterosemy have been understood.

Section 6: More generally, this section could be more elaborate. The stance of the author, formulated on p.23 is clear enough, but the theoretical stakes are high here, and deserve to be laid out in more detail and with fuller referencing. One thing that is not mentioned at all is the option that a lexeme does *not*change its core semantics when used in a pragmatic/syntactic function that it is not prototypically associated with, as is arguably the case for predicated object-denoting lexemes and for referentially used action-denoting lexemes.

What was previously Section 6, as mentioned, is now a longer discussion of semantics under Criteria 6 and 7. This is also supported by a more solid definition of the lexeme established in Section 3.2 that helps undergird the discussion of semantics.

**Reviewer 2**

This is a very interesting and clearly-written paper and I heartily recommend publication, subject to some minor revisions set out below. It brings together two general methodological approaches, arguing that they allow us to give a more useful and revealing treatment of the rather deadlocked ‘flexible category’  issue in word-class typology:

1. the use of canonical typology to establish clear reference points against which particular languages can be measured, even where they fail to fully satisfy all standard criteria (rather like the role of cardinal vowels in IPA as opposed to language-particular instantiations). This approach has of course seen a number of recent publications by Grev Corbett and his colleagues. But given the issue of how to establish cross-linguistic comparability is still so hotly debated (with Haspelmath and Bickel, for example, both adopting somewhat different positions on this) that I found the exposition of the Canonical Typology position given here very enlightening and useful as a clear theoretical restatement of this approach.

I've tried to structure the paper so that it accomplishes this last point even better now.

1. the increasing trend to analyse linguistic phenomena (here: word-class membership) in stochastic or gradient rather than absolute terms. Here I would have found it useful to bring in a few more historical references to the origins of more gradient approaches to work by Ross on squishy categories and Rosch, Geeraerts and others on gradient categories

The discussion of Rosch and approaches in prototype theory and cognitive linguistics is now quite extensive, and found mostly in the introduction.

The overall effect is to point the way to a much clearer definition of word-class flexibility, while still observing the rigorous criteria for demonstrating word-class merger set out in Evans & Osada (2005a,b).

However, a key analytic point remain unresolved and I think it is important they be addressed in revising the paper. This pertains to the notion of lexeme, which lurks implicitly through the paper without ever being defined. If we take lexemes to be three-part signs (à la Melcukian sign, also employed in GPSG, etc), which augment the Saussurean sign by adding a combinatoric element, then the sort of situations being described sometimes involve underspecified combinatorics, with more more or less variation in meaning, and sometimes involve a specified combinatoric increment accompanied by some combinatoric shift (Lichtenberk’s notion of heterosemy is useful here), and sometimes hold the form and combinatorics constant (at least at some level of granularity) while varying the meaning (what is normally treated as polysemy). Lexicographers have adopted useful principles, and metalanguage, for dealing with these issues – with lexeme and lemma being the most important ones. I think the author needs to state their position on this quite early in the paper – and of course think through which aspects of their analysis it impacts upon, and adjust them as necessary – because without them it is not clear which situations their analysis is applying to. My guess is that the author will say: ‘all’ (perhaps excluding accidental homophony, but then criteria are necessary), and if this is the case then the way these different situations are distributed across the non-canonical space needs to be made clear. (As does the fact that lexicographers employ rather different principles among themselves; I think Melcuk’s team’s explanatory combinatorial dictionaries of Russian and French are useful places to start because of the rigorous statement of principles employed).

As mentioned above, this comment was the basis for a major restructuring of the paper. I do adopt Mel'čuk's three-part sign now, but I revise the notion so that a sign allows for flexibility along any of its three dimensions: semantics, syntactics, and form. I attempt to make the case for defining the lexeme this way, so as to allow for the possibility of flexibility, rather than placing ad hoc limits on how far a sign is allowed to vary. All this is the central topic of Section 3.2 now.

[All of this reviewer's other comments were very minor, just requiring clarification, rewording, or in a few cases a quick explanation. Many of them became irrelevant as I thought more about the lexeme and integrated the discussion into my paper, making clarifications. I've responded to a few of the comments below, but left out a number of the other ones. Just let me know if there are any you'd like me to respond specifically to that I didn't.]

p. 8, 3-4th last line of 2nd para. ‘to the extent that .. may switch’ > ‘if we accept the premise that its lexemes may switch’

The definition of the lexeme given in Section 3.2 now gives an argument for why we should accept this premise.

§4.2. defn of Flexible Category 2. Here I really felt at sea without meaning being brought into the equation. It’s not the same as function.

Addressed in the larger discussions of semantics now.

p. 10, mid. ‘truly and confoundingly ambiguous’. This seems naive – pragmatics can do a great deal, and I think it’s unwise to treat it as part of the linguistic system per se since pragmatic principles apply to so many other types of communication a well.

An excellent point. I've been more careful to treat pragmatics as separate from the linguistic system per se, but it's also important to acknowledge Reviewer 1's point about the limitations of pragmatic interpretation. While I tend to agree with the present reviewer that you can get a lot from pragmatics, under Criterion 7 I work under the assumption that pragmatics is limited in what it allows one to infer, thus taking Evans & Osada's critiques of lexical flexibility seriously, and show how a canonical approach deals with this problem.

p. 12, comments on Iñupiaq after (7). I’d say the opposte: i.e. it compensates for the rigidity of its word classes by having a rich set of derivational affixes

An interesting point. I acknowledge this in the paper after example 19 now.

p. 16, 1st para after Criterion 6. But note that, in Mundari, not all implement names exhibit the same semantic pattern (examples are given in the original article of both ‘use as an X’ and ‘make an X’ depending on the lexeme)

I now go through and address these different semantic patterns in depth under Criterion 7

p. 19, top. Worth noting that, at least on Himmelmann’s analysis, Tagalog works by having more noun-like predicates which are ‘oriented’ by what have variously been called voice, focus etc. affixation; of course the whole Tagalog question remains widely debated but Himmelmann’s analysis stands out as one that takes predicates in Tagalog to be more noun-like than verb-like in their organisation.

Very helpful comment. I've used Tagalog as an example for Criterion 10.

p 22, (24).   Finally something like a set of definitions for lexeme and other categories is given, but the terminology is odd. (b) looks like like a case of heterosemy, if I’ve understood correctly. With (c) I’m not sure if this is still allowing for similar combinatorics, which would effect whether they should really be treeated as separate lexical entries. Calling (d) ‘polysemy’ seems odd to me – it appears to then restrict polysemy to cases where the semantic relationship is diachronic. This would rule out e.g. well-known polysemies like those of ‘see’ in English as cases of polysemy.

Addressed in Section 3.2.

Para under (24) – once again, one feels a need for some discussion from a lexicological perspective. Spencer’s position is close to that held by those producing dictionaries in the Melcukian tradition, which are based on rigorous modelling of the  combinatorics. But that tradition – among others – also feels the need for other, more inclusive levels (like the lemma) to capture the insights of relatedness between lexical entries that may differ in e.g. their combinatorics.

This also refers to the list of different approaches to semantic shift that Reviewer 1 had concerns with. The discussion here should be much clearer now.

**Reviewer 3**

[A general note regarding this reviewer's comments: Comments from the previous reviewers have helped me clarify the relevant concepts, especially that of the lexeme, in a way that I think now addresses many of this reviewer's questions and criticisms. Laying out my framework and theoretical positions more precisely in the introductory sections should resolve many of these. And I've noted above some ways in which I've addressed this reviewer's concerns throughout the paper, especially by situating the paper in the literature more.]

I am afraid that I cannot recommend publication of this MS. The MS presents a method for the typological classification of languages, with respect to certain properties of word classes, using the methodology of canonical typology. However, it does not actually show a classification of any particular language or set of languages along the typology proposed. Nor does it show how this typological classification is an improvement on the others in the literature, in particular those of Hengeveld and his colleagues, and of Croft (which are very different from each other). For example, it does not show how the typological classification would capture crosslinguistic generalizations of the sort that Hengeveld or Croft claim to capture. Finally, the classification itself is vaguely defined.

An application of the typology is given for Mundari in the conclusion. In the introductory sections, I show the specific problems with Hengeveld's concept of lexical flexibility, apply critiques from the literature, and revise it.

The concept of a flexible word class is associated with the typology of Hengeveld and his colleagues; the concept is rejected, or perhaps simply treated as irrelevant by Croft, except for the issue of semantic shift (see the discussion of Criterion 6 below). As such, the author needs to situate her/his typology in the context of the current issues in the area. The author should present the Hengeveld typology of flexible/rigid languages to the reader, and then show its flaws and eventually how the author's canonical typology approach is an improvement on the Hengeveld typology. The author should also present Croft's critique of the Hengeveld typology (in Croft 2001, cited by the author) and Croft's alternative approach, and either show how the author's definition of flexible/rigid incorporates Croft's criticisms of Hengeveld's typology, or defend the notion of a flexible/rigid category contrast as a useful comparative concept (in the sense of Haspelmath 2010) against Croft's criticisms. While there are issues of theoretical consistency that need to be addressed, such as Croft's charge of methodological opportunism in word class definitions, the most effective argument for the utility of the author's typology would be to show that it captures some interesting typological universals that the other approaches do not capture. I found the actual typological classification proposed by the author to be vague. I will briefly mention my concerns by criterion.

The structure of the paper now reflects these suggestions, as noted under the previous comment, and my comments at the beginning of this document.

Criterion 1: presence/absence of function-changing derivational morphology. This criterion is used both by Hengeveld (1992) and by Croft (2001; he calls it structural coding); the author should point this out. However, most of the examples that the author gives involve changes in semantic type (object, property, action), not just changes solely in function (e.g. action nominalization, nominal predication) - unlike Croft (2001) or even Hengeveld (1992), who is not consistent about this. One might expect that a change in semantic type will also involve a change in word class, as Croft argues. Also, the scope is not defined: changes from which function to which function? Do all words in the class require/not require derivational morphology?

I make the connection to Hengeveld and Croft explicit now. A much wider variety of examples are now featured in the paper.

Criterion 2: presence/absence of inflectional morphology. Again, how much morphology, and for which categories? Some languages have rich nominal morphology and little verbal morphology, others have the other way around, still others have lots (or little) of both. And that is not to mention morphology associated with adjectival modification, numeral modification, etc.

The criterion is meant to be general enough to encompass all these possibilities. I now address this, and describe how you would assess flexibility in some of these cases.

Criterion 3: all/some words in a class 'have the exact same distribution'. Distribution as defined by what constructions? There are a very large number of constructions in a language; no principles are given to explain which ones are to be used to evaluate this criterion. Note that in construction grammar, used by Croft, morphological inflections are constructions, and Hengeveld (1992 at least) doesn't strongly distinguish between morphological and syntactic means to express functions; so Criterion 3 can't be distinguished from Criterion 2 (or 1).

I think it's fairly clear from the initial discussion that *all* constructions are on the table when it comes to determining distribution, much in line with Croft's Radical Construction approach. I've tried to make this more explicit here now as well. It also shouldn't bother us that it is occasionally difficult to draw discrete lines between criteria, since the criteria converge on the same point. As to the final point, even though the dividing line between morphology and syntax is often fuzzy, and both are types of distributional potential, that doesn't mean we shouldn't distinguish them. Croft himself distinguishes between two types of distributional potential - inflectional potential (the inflectional values a lexeme takes) and distributional potential (the syntactics, or constructions it appears in). I see no reason why a similar distinction should not be made here. The point of Criterion 3 is really that, in isolating languages the inflectional potential of every lexeme is the same, so the language is more flexible.

Criterion 4: "maximum" vs. "restricted" combinatorics. It is hard to tell how this criterion is distinguished from the others - by "combinatorics", the author means distributional potential. The author cites Evans and Osada 2005a, but gives no explanation; an explanation must be provided.

Criteria 3 and 4 are very similar, so I spend a few paragraphs clarifying their differences now.

Criterion 5: feature values apply to all/some lexical items. It appears that this criterion is closest to Croft's behavioral potential. Croft proposes specific universals about behavioral potential and structural coding (and Hengeveld 1992, who looks only at the latter, also proposes similar universals). Does the author's typology allow her/him to capture the same universals, or improve on them?

Specifically, this is similar to a subtype of behavioral potential that Croft calls inflectional potential. But where Croft's criterion is interested in the markedness of non-prototypical uses of a lexeme, my Criterion 5 is interested in the grammatical features (grammemes) that a given lexeme is allowed to take. The two make use of the same concepts, but are otherwise unrelated to each other. My criterion is an assessment of which values a lexeme may take compared to other lexemes, whereas Croft's is an assessment of the feature values of a lexeme in one function as compared to another function. I think perhaps the reviewer has simply misunderstood this criterion; it's fairly tangential to Croft's. But I've still tried to clarify it better in the paper.

Criterion 7: predictable/unpredictable semantics. Predictability is a strict criterion. The author seems not to treat it very strictly (see section 6 of the MS, where there is no real discussion of the criteria for vagueness vs. polysemy vs. homonymy - a huge issue in semantics). This is a place where Hengeveld's and Croft's theories diverge dramatically. As far as I can see, this criterion cannot capture Croft's (2001) universal that zero-coded semantic shifts are always towards the semantic class prototypically expressed by the relevant function.

Issues of semantic shift are now handled much more extensively in the paper, particularly in relation to the definition of the lexeme. I've also now suggested a way in which Criterion 6 (now 7) is at least commensurate with Croft's universal regarding semantic shift, although the reviewer is right that this criterion cannot derive Croft's universal itself. But I don't think the criterion should need to. It attempts to assess the extent of the semantic shift in zero-conversion; the direction of the semantic shift is somewhat irrelevant (although it does raise an interesting implication, as I now note under this section).

Criterion 8: multiple/just one prototypical category/function. How many functions are defined? Just the "big three" (predication, reference, modification)? The author distinguishes adverbial modification. What about quantification or deixis? Also, the particular example given, from Tonkawa ('top' [N] alternates with 'up' [Adv]) is hardly general - most nouns cannot easily be used as directional adverbs.

I've structured the paper to more consistently focus on the 'big three' now, but I do note in the introductory sections now that these are just the most central examples, and I give examples of what some others might be. I also provide an alternative formulation of this criterion, which simply states that 'lexemes which occur in many constructions are more flexible than lexemes which do not'. This shifts the criterion from a focus on functions to distributions more broadly. As to Tonkawa, the point is to show an instance of a flexible *lexeme* according to Criterion 7, so this isn't problematic. Criterion 9 deals with the problem of generality.

Criterion 8: mono-/multi-categorical language. Here, it is unclear what the author means. There is a consistent conflation in the MS of functions (predication, reference), which are expressed by constructions, and word class (verb, noun), which are properties of words. It is also unclear to me how this is different from Criterion 7.

I actually think the monocategorical/multicategorical distinction is fairly straightforward. However, the reviewer is right that I needed to be clearer about what I meant by functions, and I have done so throughout now. The reviewer is correct that this criterion is a special case of the previous one, but it is still worth discussing separately. It may be that some lexemes in a language are omnifunctional, and these lexemes are therefore flexible according to criterion 8, but only if these lexemes pervade the lexicon would we want to call the language monocategorial. So the two criteria are logically independent of each other.

Criterion 9: every/some lexeme is flexible. Again, based on the preceding criteria, there are many different ways in which a lexeme could be flexible. Criterion 9 would have to be relativized to all the values of most of the preceding criteria (once they are more precisely defined).

This is a helpful insight. A lexeme can really only be considered more-or-less flexible in the aggregate, considering all the ways it might be flexible. Alternatively, one could assess a language according to Criterion 9 for each of the previous criteria outlined - tedious, to be sure, but not in principle impossible. I've made this clarification now.

Criterion 10: open/closed class. Again, this depends on which constructions/distributions are used to define word classes. So it can only make sense when relativized to which construction(s) are used to define word classes.

I've decided to remove this criterion from the paper, in part because of the reviewer's comment, in part because I'm no longer convinced it's true even when properly formulated.

How much is this due to the nature of canonical typology? I don't really know. The author's description of canonical typology makes it look like typological prototypes: the most canonical instance is the most prototypical member of the category across languages. That is a reasonable approach to typology. But what matters is finding the combination of feature values that defines a typologically significant prototype, that is, one for which a typologist can construct interesting language universals. It is not clear to me that there are significant language universals that are associated with the flexible category prototype that the author aims to define in the MS. In fact, the author suggests that the prototype isn't even attested, and may not even be possible, as a human language.

I spend some time discussing the differences and overlap between canonical types and prototypes now. I also explain why the non-existence of canonically flexible categories is nonproblematic. The point is to assess languages in *relation to* this canonical type, regardless of whether this canonical type exists - it serves as a standard of comparison in a way different than prototypes do. So I do think a number of this reviewer's comments are simply due to a difference in understanding regarding the method of canonical typology. Accordingly, I explain the concepts of CT much more thoroughly and explicitly now, and continue to make clarifications throughout the paper, with many more examples to illustrate. Finally, this article does not attempt to draw significant language universals, but rather to present a typology of lexical flexibility. The reviewer was right that the earlier draft didn't do a good job actually typologizing languages according to this typology - I just put the framework out there and didn't really do much with it. But now, through more examples, and through an explicit first attempt to typologize Mundari according to its lexical flexibility in the conclusion, the paper does a much better job at this. I think this reviewer will better understand how the present typology can be applied to languages after reading this version.

**Reviewer 4**

Overall, it is a good summary and critique of rigid vs flexible categories. However, I think section 2 could be elaborated especially for the benefit of those unfamiliar with canonical typology, with real language examples thrown in. In fact, the paper could do with more examples throughout. Also, it could be improved on by bringing Cognitive Linguistics’ take on the issue (e.g. prototype) into the picture. While one paper by Langacker is listed in the references, there is hardly any discussion as to how canonical typology stands vis-à-vis CL on the issue. I understand that the author is concerned mainly with linguistic typologists' handling of the issue but it will be remiss of him not to include CL in his discussion. This seems to me to be a serious gap. Sections 2 and 6, in particular, will benefit greatly from this addition.

All of this reviewer's suggestions have been taken gratefully. As mentioned, the section on Canonical Typology is more detailed now, explicitly discussing its relationship to prototype theories. I've also included many more language examples to illustrate the spectrum of possibilities for each criterion.