Separating Semantic and Morphosyntactic Clusters: A Note on Transparent Free Relatives

Brett Reynolds

Humber Polytechnic & University of Toronto

Draft for discussion – January 2026

Abstract

This note proposes a diagnostic refinement for TFR analysis: explicitly distinguishing morphosyntactic constraints (which project distributional generalizations) from semantic/pragmatic constraints (which project interpretive generalizations). The SBCG formalism already separates these in the typed feature geometry; my concern is that several *informal claims and diagnostic arguments* in the current analysis slide between them. Using the Homeostatic Property Cluster (HPC) framework, I treat these as distinct-but-coupled natural kinds, each stabilized by the other. The payoffs are sharper diagnostic arguments, a refined treatment of conventional implicature as source attribution rather than hard-coded non-commitment, and a concrete programme of dissociation tests. A companion note applies the same framework to independent relative *whose*.

1 Introduction: The diagnostic hygiene problem

The core methodological claim of this note: don't let a single label quietly do two jobs.

Your TFR analysis uses terms like "definiteness", "transparency", and "nucleus" in ways that sometimes slide between morphosyntactic and semantic readings. In most cases this doesn't matter – the two profiles align. But at the boundaries, the conflation obscures which profile is doing the explanatory work, and it closes off empirical questions that the separation would open.

Consider existential constructions. In *There is what they call a foreigner at the door*, the TFR is grammatical. Swap the nucleus: ?*There is what they call the foreigner at the door*. The sentence degrades. But what constraint is violated? Not morphosyntactic deitality alone – definite *there*-sentences exist with presentational or list readings (*There's the guy I told you about*). Not semantic definiteness alone – the discourse-newness requirement of canonical existentials isn't identical to either. We may need three notions here: morphological deital marking, semantic definiteness (unique identifiability), and discourse-status licensing for existentials. A single label papers over the distinctions.

This note proposes separating the profiles explicitly. Your formal SBCG machinery already distinguishes SYN and SEM in the typed feature geometry – I'm not suggesting architectural changes. My target is the informal prose and choice of diagnostics: places where the analysis treats coordination, *there*-licensing, or agreement as if they diagnose a unified "category" when they actually probe different profiles.

The framework I'll use is Boyd's Homeostatic Property Cluster theory (Boyd, 1999). The HPC vocabulary is just a way of regimenting a familiar practice: separate the properties that support distributional generalizations from those that support interpretive ones, then ask where they come apart. Morphosyntactic and semantic categories are distinct HPCs – stable configurations of co-occurring properties maintained by causal mechanisms. Each includes the other as a stabilizing mechanism: semantics helps maintain morphosyntactic coherence, and vice versa. But they remain distinct because they *project* for different purposes: the syntactician's category supports distributional generalizations; the semanticist's supports meaning generalizations.

The payoff isn't philosophical elegance; it's a concrete research programme. Bundled constraints don't just explain rarity; they predict *where to look for counterexamples*. Counterexamples aren't nuisances – they're the data that tell you which part of the bundle is doing which work.

2   The framework: two coupled HPCs

2.1   Field-relative projectibility

Consider *Paris*. A syntactician classifies this as a PROPER NOUN – a morphosyntactic category defined by distributional properties (articleless in English, capitalized, resists pluralization in core uses). The syntactician is aware of the typical semantic properties of this category – it's a rigid designator, it designates a unique referent, etc. – but he considers them to be only a small part of the definition. A semanticist classifies the same expression as a PROPER NAME – with the semantic ideas that entails. The semanticist is aware of the typical morphosyntactic properties of this category – it's articleless, capitalized, etc. – but they don't occupy his attention.

These aren't the same category. They overlap massively in extension, but they have different membership criteria (distributional vs. meaning-based), support different generalizations (syntactic distribution vs. truth conditions), and come apart at the edges (nicknames, converted common nouns, etc.).

Following Goodman's notion of projectibility (Goodman, 1955), each category is *stable for a purpose* – it supports reliable inductive inference within a domain. What makes a property cluster "project well" for syntactic purposes is that the homeostatic mechanisms maintaining it are sensitive to distributional regularities; what makes it project well for semantic purposes is that the mechanisms are sensitive to inferential and truth-conditional regularities. The syntactician's "proper noun" and the semanticist's "proper name" are both legitimate natural kinds, but they are *field-relative*: natural for different fields of inquiry.

2.2   The HPC model

The Homeostatic Property Cluster framework, developed in philosophy of biology to handle "messy" natural kinds like species (Boyd, 1991, 1999), provides a formal way to characterize these categories. An HPC kind is defined by:
1. a cluster of co-occurring properties (no single property is necessary or sufficient),
2. homeostatic mechanisms that causally maintain the clustering, and
3. fuzzy boundaries and graded membership as a predicted consequence.

Applied to linguistics:

Morphosyntactic HPC.   A cluster of formal and distributional properties: morphological marking, syntactic position, agreement behavior, selection restrictions. The cluster is maintained by mechanisms including syntactic licensing, paradigmatic pressure, and – crucially – correlation with prototypical semantic functions.

Semantic HPC.   A cluster of meaning and function properties: identifiability, familiarity, referent type, discourse status. The cluster is maintained by mechanisms including conceptual coherence, inferential utility, and – crucially – correlation with prototypical formal expression.

2.3   Mutual homeostasis

Each cluster can plausibly include the other as a stabilizer: recurrent form–meaning alignments make both the formal class and the interpretive function easier to learn and to maintain over time (in acquisition and in diachrony).

The consequence: form and meaning will typically align, but the alignment is imperfect because the reinforcement isn't deterministic. Where form and meaning pull in different directions – weak definites, generics, proper names – the clusters diverge. The divergence isn't anomaly; it's the expected outcome of imperfect homeostasis.

2.4   A reusable comparison

Table 1 summarizes the distinction in a form that can be applied across constructions.

Dimension	Morphosyntactic category	Semantic category
Profile purpose	Projects combinatorics and distribution	Projects interpretive and inferential behavior
Stabilizers	Agreement, selection, constituency, positional regularities	Truth-conditional roles, CI/pragmatics, type-shifting patterns
Primary diagnostics	Coordination, agreement, selectional fit, island sensitivity	Entailments, scope, anaphora resolution, felicity conditions
Cross-domain supports	Semantic regularities can cue syntactic category	Morphosyntactic patterns can cue semantic type
Failure mode if collapsed	Syntactic diagnostics overdetermine meaning	Semantic expectations overdetermine form

Table 1: Comparison of morphosyntactic and semantic categories.

3   Case study: transparent free relatives

3.1   What the analysis gets right

Your construction-based analysis captures several key properties elegantly:

1. Parenthetical function: The TFR clause (*what they call*) is omissible without changing truth conditions.

2. Nucleus determines morphosyntactic properties: The predicative nucleus (*a foreigner*) controls agreement and category, and tracks distributional effects relevant to deitality/*there*-licensing (with semantic definiteness and discourse status as separable factors).
3. Constructional constraints: Two interacting constructions – the Transparent Free Relative Clause Construction and the TFR Construction – license the structure.
4. Projective meaning: The construction contributes non-at-issue content commonly described in terms of non-speaker commitment; below I argue that the most stable generalization is *source attribution / metalinguistic qualification*, with commitment effects derived pragmatically.

The analysis places the nucleus as a sister to the free relative clause (rather than internal to it), which cleanly captures extraction possibilities, binding facts, and idiom integrity. The SBCG formalization is precise and empirically grounded.

3.2   WHERE DIAGNOSTICS SLIDE BETWEEN PROFILES

Despite these strengths, several informal claims treat semantic and morphosyntactic categories as a unified profile, letting one do the work of both. These are framing compressions, not data errors, but they blur which profile is doing the explanatory work. I identify six representative points, with suggested rephrasings.

1. Nucleus determination bundles distinct profiles.   The wh-phrase in SFRs is described as the "syntactic and semantic nucleus", and the TFR nucleus is said to "determine the syntactic category and semantic properties of the entire construction". This treats two HPCs as one.
   Suggested: "morphosyntactic nucleus; semantic profile is constructionally composed" (i.e., nucleus meaning + constructional CI + pragmatic licensing).

2. Coordination diagnoses functional parallelism, not category.   Coordination is treated as a diagnostic for category identity. But coordination is sensitive to multiple constraints: functional parallelism, syntactic category, and semantic compatibility. These don't always align. So coordination cannot by itself settle which *profile* is inherited – it shows functional parallelism; category and semantic type require independent argumentation.
   Suggested: Treat coordination as evidence for functional parallelism; note that category match and semantic compatibility are independently motivated.

3. Existential licensing is discourse-status, not semantic definiteness.   "The nucleus's definiteness determines grammaticality" in existentials. This folds semantic definiteness (identifiability, uniqueness, familiarity) into morphosyntactic deitality (*there*-resistance, form-class membership) into discourse-status licensing (presentational/new-information requirement). The three can diverge: weak definites are deital but indefinite-like semantically; definite *there*-sentences exist with presentational readings. (Discourse-status licensing is part of the semantic/pragmatic profile – or interface profile – but crucially does not reduce to semantic definiteness or morphosyntactic deitality.)
   Suggested: "The nucleus's morphosyntactic and discourse-status properties control *there*-licensing; semantic definiteness patterns may align but are independently motivated."

4. Underspecification has to distinguish syntactic from semantic resolution.   *What* is described as "categorially and semantically underspecified". The wording bundles syntactic underspecification with semantic type underspecification. Syntactic underspecification allows *what* to unify with any predicative category; semantic type resolution is a separate process driven by the nucleus and the constructional CI.

   Suggested: "syntactically underspecified; semantic type resolved by the predicative nucleus and construction".

5. Nucleus determines morphosyntax, not semantic projection.   "The nucleus determines phrase-level properties." This is morphosyntactic headship, but the phrasing invites a semantic-projection reading. Semantic effects come from the construction plus nucleus, not from the nucleus alone.

   Suggested: "the nucleus determines morphosyntactic properties; semantic projection is licensed constructionally".

6. Verb inventory reflects coupled constraints.   The paper identifies verbs that occur in TFRs as those selecting predicative complements and expressing non-speaker commitment. This is naturally analyzed as two coupled constraints – one morphosyntactic (predicative selection), one semantic (epistemic hedging). The separation predicts a 2×2 matrix of possibilities:

   1. Selects predicative, allows hedging (*call, consider*): Canonical TFR verbs.
   2. Selects predicative, no hedging (*be, become*): Prediction: Should fail if the CI is obligatory, or require a different constructional meaning.
   3. No predicative selection, allows hedging (*guess, suspect*): Prediction: Impossible in TFR syntax despite semantic compatibility.
   4. Neither: Impossible.

These distinctions make the "coupled constraints" claim concrete and falsifiable.

4   THE CI STORY NEEDS REFINEMENT

4.1   THE PROBLEM WITH HARD-CODED NON-COMMITMENT

Your analysis treats non-speaker-commitment as a stable constructional meaning. But the dissociation programme forces a harder question: does the CI survive when the source is the speaker?

   Consider *what I call X* vs. *what they call X*. When the speaker is the source, the utterance often functions not as hedging but as a naming or terminological move:

(1)   a. *This is what I call the "transparency effect".*
       b. *She's developed what I call a robust methodology.*

   In these cases, the speaker isn't distancing from the predication – they're *claiming* it, often with metalinguistic emphasis. Conversely, other-source TFRs can occur without distancing when the speaker fully endorses the label:

(2)   *She won what the committee called "Best in Show" – and rightly so.*

   If these patterns hold, then "non-speaker commitment" is too specific for the CI. A more defensible claim, consistent with work on projective content (cf. Potts, 2005; Simons et al., 2010):

The construction contributes projective, non-at-issue content best characterized as source attribution / metalinguistic qualification: who applies the predicate, or under what evidential standard. Speaker-commitment effects are a downstream pragmatic computation that depends on source, evidential verb, and discourse goals.

This is a substantive refinement, not a minor rewording. It predicts that TFRs with different source configurations will show different commitment profiles – testable via embedding and challenge tests.

4.2   A TEST PROGRAMME FOR THE CI

If the CI is source attribution rather than hard-coded non-commitment, we should find systematic variation along these dimensions:

1. Speaker-source vs. other-source: *what I call X* vs. *what they call X*. Prediction: speaker-source licenses endorsement readings; other-source licenses distancing readings; the construction itself is neutral.
2. Evidential gradient: *what is said/reported/rumoured/known/claimed to be X*. Prediction: evidential strength modulates speaker commitment independently of source.
3. Predicative complement types: AP vs. NP vs. PP vs. VP/gerundive (*what we call picking corn* is in your data). Prediction: CI stability across complement types; if it varies, that's evidence for complement-specific effects.
4. Embedding beyond negation/conditionals: questions (*Is breast cancer what scientists call multi-factorial?*), imperatives (*Be what they call professional!*), metalinguistic negation. Prediction: CI projects in all environments if it's truly conventional; if it fails to project somewhere, that constrains the analysis.
5. Prosodic integration: comma-parenthetical vs. prosodically integrated. Prediction: prosodic separation may strengthen the metalinguistic/hedging reading; integration may weaken it.

5   TWO KINDS OF TRANSPARENCY

Once we separate the clusters, we can distinguish two senses in which TFRs are "transparent":

Morphosyntactic transparency.   Formal features of the nucleus percolate to determine properties of the whole:
– Agreement: [*What appear to be trousers*] *are really leggings.*
– Category selection: nucleus AP → whole functions as AP.
– Deital licensing: indefinite nucleus → compatible with existentials.

This is feature unification in the syntax.

Semantic transparency.   The nucleus contributes directly to truth conditions:
– Referent: the TFR refers to whatever the nucleus denotes.
– Proposition: the matrix predicate applies to nucleus meaning.
– CI: the TFR clause adds source attribution but not truth-conditional content.

This is compositional semantics.

5.1   Why they typically align

In TFRs, morphosyntactic and semantic transparency go together. This isn't coincidence but the result of coupling mechanisms:

1. Parenthetical syntax creates a structural configuration where the clause is "backgrounded", reducing its contribution to both formal features and truth conditions.
2. Verb semantics (*seem, call, consider*) inherently express perspective or evidentiality, mediating between the proposition and speaker stance.
3. Prosody (parenthetical intonation, optional comma setting) reinforces the non-at-issue status both syntactically and semantically.

These mechanisms explain why the two kinds of transparency cluster – and predict that they might come apart if the mechanisms are disrupted.

5.2   The dissociation prediction

If SYN and SEM transparency are genuinely distinct, we should expect to find edge cases where one succeeds without the other:

– Morphosyntactic transparency without semantic transparency: Features percolate but truth conditions don't follow the nucleus. Possible test case: TFRs with intensional verbs like *what distinctively marks* or *what theoretically constitutes*, where the nucleus's referent is world-dependent.
– Semantic transparency without morphosyntactic transparency: Nucleus contributes to truth conditions but features don't unify. Possible test case: category mismatches like **He is what they consider nice* (where *nice* is semantically appropriate but fails the NP-gap requirement of *what*).

These aren't rhetorical gestures. They're predictions. If we find them, the separation earns its keep; if we don't, we learn that the coupling is tighter than the framework suggests.

6   Revised SBCG analysis sketch

6.1   The linking constraint as the key addition

Your formal analysis already separates SYN and SEM in the typed feature geometry. What the HPC framework adds is a way to make the *linking constraint* – how the predicative-complement semantics is compositionally tied to the [PRD +] configuration – explicit and central.

The construction should encode:

– A SYN constraint: gap + [PRD +] + head inheritance from nucleus.
– A SEM constraint: property-ascription + attribution frame.
– A linking constraint: the predicative-complement semantics is licensed by the [PRD +] configuration; the attribution CI is licensed by the verb's evidential/perspective semantics.

This is the SBCG-friendly version of "mutual homeostasis": the linking constraint states how the syntactic and semantic requirements are causally connected, not just co-present. I'm not attempting a full alternative formalization here; the point is just to locate the coupling claim in an explicit form–meaning linkage (rather than in informal diagnostics that slide between profiles).

6.2   THE TRANSPARENT FREE RELATIVE CLAUSE CONSTRUCTION

The first construction combines *what* with an incomplete S:

$$\textit{Trans-Free-Rel-Cl} \rightarrow \boxed{1}\,\text{NP[FORM } \textit{what}\text{]}, \text{S[VFORM fin, GAP } \langle\boxed{1}\,\text{NP}, \boxed{2}\,\text{XP[PRD +]}\rangle\text{]}$$

The resulting S has category *tran-free-rel*, retains $\boxed{2}$ XP on its GAP list, and introduces a SEM.CI value encoding source attribution for the embedded predication. The key point: SYN.GAP specifies what has to be filled; SEM.CI specifies how the filling is interpreted.

6.3   THE TFR CONSTRUCTION

The second construction combines the transparent relative clause with its nucleus:

$$\textit{TFR-Cxt} \rightarrow \text{S[}\textit{tran-free-rel}\text{, GAP } \langle\boxed{1}\,\text{XP}\rangle\text{]}, \text{H}\,\boxed{1}\,\text{XP[PRD +]}$$

The result inherits SYN features from the nucleus (the H daughter):
– SYN.CAT is identified with the nucleus category (morphosyntactic transparency).
– SEM.INDEX is the nucleus referent (semantic transparency).
– SEM.CI encodes source attribution for the predication.

The linking constraint is implicit in the construction's form-meaning pairing: the [PRD +] requirement on the gap corresponds to the property-ascription semantics; the verb's selection of a predicative complement corresponds to the attribution frame.

7   CONCLUSION

The proposal: treat morphosyntactic and semantic categories as distinct-but-coupled HPCs; adjust informal diagnostic claims to respect the distinction; refine the CI as source attribution rather than hard-coded non-commitment.

The CI revision is where the separation pays off most. If the projective meaning is source attribution (in the family of evidential and perspective-marking constructions), then commitment is a pragmatic downstream effect that should vary predictably with source, evidential strength, and discourse goals. That prediction is testable.

The payoffs are precision, mechanism, and testability. Precision: we can state constraints on form separately from constraints on meaning, and make their interaction explicit. Mechanism: the coupling between parenthetical syntax and attribution semantics is explained by homeostatic forces, not stipulated. Testability: the framework predicts dissociations and provides a programme for finding them.

At the deepest level, both TFRs and independent relative *whose* (the subject of the companion note) are about *underspecification and resolution*. In *whose*, the possessum type is underspecified and has to be resolved from discourse. In TFRs, the attribution frame – and thus the commitment profile – is underspecified and resolved from source, evidential verb, and discourse function. The HPC framework makes this parallel visible: both constructions have a structurally guaranteed component and a discourse-dependent component, and the discourse-dependent component is where the action is.

The broader methodological point: bundled constraints predict where to look for counterexamples. Counterexamples are not nuisances – they're the data that tell you which part of the bundle is doing

which work. Trusting the bundle leads us to mistake the silence of the corpus for the prohibition of the grammar. We conflate what is merely *improbable* (the convergence of separate licensing conditions) with what is structurally *impossible*.

AI Acknowledgment

I acknowledge the use of Claude Opus 4.5, ChatGPT 5.2, and Gemini 3 in the drafting of this paper. I take full responsibility for the content.

References

Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies, 61*, 127–148.

Boyd, R. (1999). Homeostasis, species, and higher taxa. In R. A. Wilson (Ed.), *Species: New interdisciplinary essays* (pp. 141–185). MIT Press.

Goodman, N. (1955). *Fact, fiction, and forecast*. Harvard University Press.

Potts, C. (2005). *The logic of conventional implicatures*. Oxford University Press.

Simons, M., Tonhauser, J., Beaver, D., & Roberts, C. (2010). What projects and why. *Proceedings of SALT 20*, 309–327.