



A Discourse Perspective on Test Item Development

James J. Bauman, Center for Applied Linguistics

jim@cal.org

LTRC 2009, Denver



ONPAR Item Development Realities

Computer-based
Low language
Multimodal
Interactive
Automatically
scored



The Cold Reality: ONPAR items are expensive!

- Items on traditional tests can have rejection rates at 50% and above following field testing.
- For ONPAR, rejection rates closer to 0% are needed to assure a cost effective test.

\$64 Question:
How can we assure high
item retention rates?

Functional Linguistic Organization of Test Items

Ideational metafunction

Functional Concerns

- Staging of the construct
- Ancillary support for construct

Functional Realizations

- 1. Signifiers and depictions of salient information**
- 2. Perspective on and windowing of information**

Interactional metafunction

- Task demand
- Response

- 3. Indicators of participant relations**
- 4. Warrants re the item**

Textual metafunction

- Item presentation mode
- Structure of response space

- 6. Representational (modal) choices**
- 7. Cohesion marking**
- 8. User interface designs**

The term “metafunction” and its classification is attributed to Michael Halliday’s theory of Systemic Functional Linguistics, (Halliday & Matthiessen, 2004).

Functional Components of the Test Item

Domain Specification

Item Perspective

Contextual Situation

Item Demand

Response Space(s)

Response

Item Writer
Contribution

Test Taker
Contribution

The Item Perspective (IP)

Merger
of:

Configurational structure describes a conceptual referent (the *figural* element) in respect to its discrete or distributed existence in space or time (the *ground* element).

Perspective describes the point of view the interactants in an information exchange adopt to the configurational structure (provides the *relational* element).

The terms 'configurational structure' and 'perspective' are taken from Talmy's (2003) taxonomy and characterization of *cognitive schemas*.

IP of Content Test Items

- In content test items, IP represents:
 - A content target (**figural element**)
 - As situated within a larger frame of reference (**ground element**)
 - From the stance the item writer expects the test taker to assume relative to the figure-ground configuration (**relational element**)

Notational Convention

Relation (**Figure**, **Ground**)

Example

Sequence (**Stages of zygote formation**, **Animal sexual reproduction**)

Sequence (**Positive & negative numbers**, **Number line**)

IP Relations

- Derivation of the IP relation notion
 - Rhetorical Structure Theory (RST)
(Mann & Thompson, 1988; Mann & Taboaba, 2006)
 - Concept of multinuclear relations: relations in which neither component can be considered secondary (satellite) to the other
- Multinuclear relations of RST

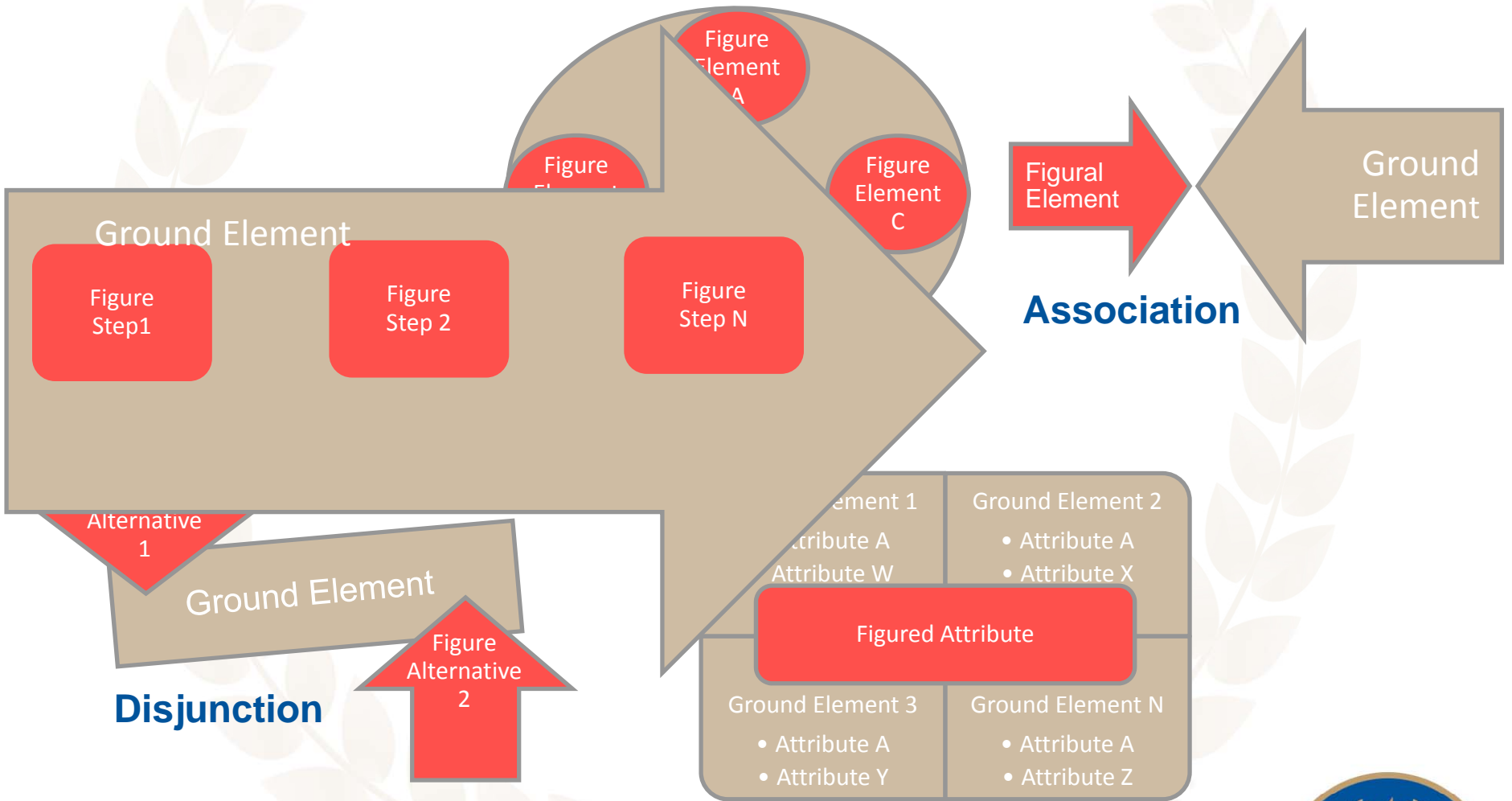
Relation Name

Cognitive Schema

- | | |
|----------------------------|---------------------------------|
| – Conjunction | Inclusion |
| – Disjunction | Substitution |
| – Sequence | Temporal order & Force dynamics |
| – Association ¹ | Organization |
| – Contrast | Property & Contingency |

¹The 'List' relation in RST is here renamed 'Association' to better convey the cognitive schema of the implied organization motivating the listed elements.

IP Relations



Examples of IP Relations

Contrast: Two or more ground elements presupposed to be similar in some attributes and dissimilar in others, one attribute of which is figured in the relation. Example:
contrast (population level, animals in an ecosystem: competitive pressures, food supply)

Conjunction: A set of figural elements related as coordinates or components to a conceptually independent ground element. Example:
conjunction (circuit components, electrical conductance test)

Disjunction: A set of figural elements related as alternatives or opponents to some conceptually independent ground element. Example:
disjunction (alternative food web components, ecosystem)

Sequence: A set of figural elements related in an ordered, temporal sequence to a conceptually independent ground element. Example:
sequence (frog developmental stages, frog life cycle)

Association: The situation of a figural element in spatial or temporal relation to an extrinsic ground element. Example:
association (Saturn, solar system)

Why Conduct Discourse Analysis on Test Items?

- To better understand item difficulty and item accessibility, especially for differentiated groups of test takers
- To facilitate the test development process, with the goal of generating more usable items more cost-effectively
- To support the principled creation of items utilizing multimodal representations
- To create professional development opportunities for teachers to scaffold instruction
- To develop a framework with high explanatory potential for describing differences between social language, general instructional language, academic language, and language of content areas

What Should a Discourse Analysis Attempt to Describe and Measure?

- Item coherence
 - Description of the extent to which discourse components, both explicit & inferred, interactively structure in the interactants the same mental representations.
 - Metric would provide an empirically derived covariate with which to analyze and interpret test performance.
- Rationale for multimodal representations
 - Description of the conceptual unity that theoretically underlies an item representation made up of components using various semiotic channels (Kress & van Leeuwen, 2001; Bateman, 2008).
 - Metric(s) used as covariates would provide a measure of relative effectiveness of different modes for different item types.

Prospective Models for Discourse Analysis

- Surface-level language models ← Language Acquisition and Psycholinguistics
- Grammatically based models ← Descriptive Linguistics and Computational Linguistics
- Pragmatically based models ← Philosophy of Language, Semantics, and Rhetoric
- Cognitively based models ← Cognitive Psychology and Cognitive Linguistics
- Social-Interaction based models ← Social Semiotics and Sociolinguistics

Caveat: This listing is much oversimplified, in that the various models have significant cross-disciplinary influences and the disciplines themselves are very much cross-pollinating.

What is an Appropriate Discourse Model for Assessment Applications?

Caveat

Other than the surface-level language models, current discourse models are not widely applied in assessment contexts.

- Model should include coverage of all systemic functional linguistics metafunctions.
- Model should characterize interface between cognitive, semantic, syntactic, & lexical phenomena.
- Model should cover recognized categories of discourse analysis.
- Model should deal with multimodal components.
- Model parameters should function independently.
- Model should be quantifiable.







Parameters of Proposed Discourse Model

- **Inferencing Load** – Balance between explicit versus inferred references for IP & Causal relations
- **Notional Transparency** – Degree to which key notions (IP & Task Demand, Situation, and Causal relations) are definitionally or contextually supported
- **Evidential Strength** – Relative degree to which the key propositions in the item can be asserted
- **Cohesion** – Degree of co-referentiality for key IP notions and consistency of IP and verbal tense & voice

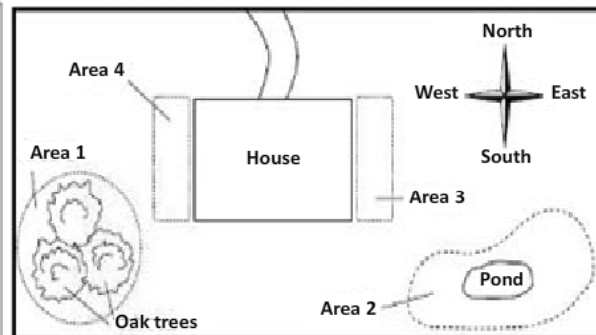
Garden Item—Derived from TIMMS

Rebecca wants to plant a garden in her yard. She studied how much sunlight different plants need in a gardening book. Look at the book and the map of Rebecca's yard below.

Gardening Book

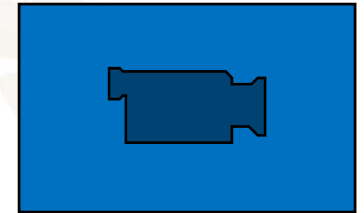
Light needed to grow well		Light needed to grow well	
Fern 	Shade	Shrub 	Afternoon Sun
Rose 	Sun All Day	Shooting Star 	Morning Sun
Wood Rush 	Part Shade	Tomato Plant 	Sun All Day

Map of Rebecca's Yard



Which correctly shows the **best** places for Rebecca to plant the flowers, shrubs, and plants in her yard?

- A Area 1- Wood Rush
Area 2- Tomato Plant and Shrub
Area 3- Rose and Shooting Star
Area 4- Fern
- B Area 1- Shooting Star and Shrub
Area 2- Fern and Rose
Area 3- Tomato Plant
Area 4- Wood Rush
- C Area 1- Fern and Wood Rush
Area 2- Rose and Tomato Plant
Area 3- Shooting Star
Area 4- Shrub
- D Area 1- Shrub
Area 2- Fern and Wood Rush
Area 3- Rose and Tomato Plant
Area 4- Shooting Star



Equivalent
ONPAR Item

Content Analysis of Garden Item

- Item Perspective
 - Contrast (sunlight availability, plant growth: sunlight requirement, location, movement of sun, barriers to sun)
- Situation
 - Four locations are specified on a yard map, situated in reference to different structures (house) or features (trees); topographic orientation is indicated by compass rose
 - Garden book with sunlight requirements for 6 plants
- Causal relations
 - Movement of sun & barriers to sunlight determine availability of sun; Sun availability determines growth

Discourse Analysis of Garden Item

- Inferencing Load
 - IP relation: Contrast relation between available sunlight & plant growth must be inferred.
 - Causal relations between sun availability and growth must be inferred.
- Notional Transparency
 - IP: Sunlight availability (not supported); plants (supported); sun requirement (supported); growth (supported); barriers (not supported)
 - Situation: Yard structures as barriers (not supported); Sun movement from east to west to reference am and pm times (not supported)
 - Causal markers: Movement of sun & barriers (not supported); availability of sun (partially supported); growth (supported)

Discourse Analysis of Garden Item (con't)

- Evidential Strength
 - Demand: Conjectural because notion "should be planted in location providing needed sunlight' is problematic
 - Response: Requires an appraisal of interactions between site location, sun movement, and sun requirements
- Cohesion
 - IP: Plants: 'plants' :: pictured exemplars; sunlight: 'sunlight hours' :: clock illus :: posted hours; location: 'where' :: pictured sites; barriers to sun - not co-referenced; movement of sun – not co-referenced
 - Tense/Voice: Consistent
 - IP: Inconsistent - item introduces a possible 2nd figural IP element in the topographic locations depicted

Garden Item Performance

- Traditional item
 - Calculated coherence measure (0-1) = .40
 - Somewhat above a chance level: $p = 0.32$ (all respondents: $N=147$)
- ONPAR item
 - Calculated coherence measure (0-1) = .35
 - Below chance level: $p = 0.05$ (all respondents: $N=375$)

Conclusions

- Applying a discourse perspective to items may help better predict item performance.
- Discourse analysis early in development may reduce item development costs and increase item retention rates.
- Discourse analysis can bridge the semiotic divide between language and non-language components in multimodal items.
- **Caveat: Work, possibly substantial, still needs to be done to optimize the discourse analytic procedures and demonstrate its reliability.**

References

- Bateman, J. (2008). *Multimodality and genre: A foundation for the systematic analysis of multimodal documents*. New York, Palgrave Macmillan.
- Halliday, M. A. K. and C. M. I. M. Matthiessen (2004). *An introduction to functional grammar*. London, Hodder Arnold.
- Kopriva, R. J. (2008). *Improving testing for English language learners*. New York, Routledge.
- Kress, G. and T. van Leeuwen (2001). *Multimodal Discourse: The modes and media of contemporary communication*. New York, Oxford University Press.
- Mann, W. C. and S. A. Thompson (1988). "Rhetorical structure theory: Toward a functional theory of text organization." *Text* 8(3): 243-281.
- Taboada, M. and W. C. Mann (2006). "Rhetorical structure theory: Looking back and moving ahead." *Discourse Studies* 8(3): 423-459.
- Talmy, L. (2003). *Toward a cognitive semantics*. Cambridge, MA, MIT press.