Tell and *Say*: Towards an OT Model of Lexical Choice

Daniel Sanford
University of New Mexico

danquixote73@yahoo.com

1. INTRODUCTION

The Natural Semantic Metalanguage approach to Semantics has a utility to lexicography that cannot be ignored. There are questions, however, which remain unanswered by a purely NSM analysis of a lexical item. The distribution, for example, of a word in discourse is not explicated by an NSM definition, nor does it always provide an explanation for the basis on which one word is chosen over another in situations where each could be seen as appropriate. Lexically simple words, moreover, are not defined because they are considered undefinable: definitions should be stated in language simpler than the word being defined, and simpler language is not available. This approach overlooks, however, disparities between primitives and their counterparts in "real" languages, and again, does not provide an explanation for distribution and choice of words.

Such questions, however, are not necessarily within the scope of an NSM analysis, and the concerns presented here should not be taken as a critique of the NSM methodology. Rather, the position taken here is that the NSM approach can be combined with other methodologies to find a utility to applications outside of its current scope. Put forward here is the idea that an NSM approach which takes advantage of discourse-analytical methods as a source of data can be used not only to precisely define lexical items, but also to answer questions as to word choice and distribution. Cumming & Ono (1997: 132) point out that "the attention that has been devoted to the lexicon has focused primarily on explanations for lexical category distinctions… rather than to choice within lexical sets." An NSM approach which takes advantage of discourse data is well-positioned to address the issue of choice within lexical sets, and a model for how

speakers do so is outlined here. Such an approach can best be understood in terms of Optimality Theory, and would therefore have the added advantage of providing a single possible framework for the understanding of semantics, phonology, and syntax.

This approach will be explored here using the English words *say* and *tell*, words similar but not identical in meaning and distribution.

2. NATURAL SEMANTIC METALANGUAGE

The Natural Semantic Metalanguage (NSM) approach, brought to the fore of semantics by Anna Wiezbicka, is grounded in two assumptions about the lexicons of human languages (Wiezbicka 1998): that every language has "an irreducible core of terms in which the speakers can understand all complex thoughts and utterances" and that "the irreducible core of all languages match… reflecting the irreducible core of human thought" (113). From these two principles are derived a method with tremendous utility to the lexicographer. Every word, despite its surface complexity, can be defined in terms of a limited set of semantic invariants (Goddard & Wierzbicka 1994), Euclidean axioms upon which more complex ideas can be understood.

The theoretical grounding of NSM theory can be (and is) debated, but the definitions produced by it have advantages independent of some of Wierzbicka's more contested claims (i.e., that the set of semantic primitives represent something innate to the human mind). Definitions composed of the 60 or so proposed semantic primitives are, of necessity, termed in words simpler than the word being defined, complex ideas explained in terms of simple ones. NSM definitions avoid, therefore, common pitfalls of lexicography such as obscurity and circularity of definition (Goddard 1998). Since, moreover, semantic primitives are understood to be universal, with equivalents in all

human languages, they should have utility in providing definitions for lexical items in any human language.

3. METHODS OF DATA ACQUISITION AND ANALYSIS

Typically, the data for NSM (or, for that matter, most any lexicographical approach) analyses of lexical items comes from isolated sentences, sometimes spoken but more frequently written, which illustrate the grammatical role of the word in the sentence and in which the meaning of the word, or aspects of the meaning of the word, are clearly profiled. Constructed sentences are also heavily relied on, the semanticist relying upon his or her own intuition as a native speaker to assess the naturalness and acceptability, or lack thereof in the case of negative evidence, of the utterance. Wierzbicka, for example, supports her definition of *blame* (1988: 153) on three examples from written language (two citations from books and one from a newspaper) and numerous examples of contrived sentences— for example, "I wouldn't blame you" is used to justify that *blame* occurs, typically, without complements.

Examples, however, from written and constructed language are not necessarily ideal. Chafe (1998: 96) writes that "ordinary conversational talk… occupies a special place as the kind of language that is most natural in both form and function, the kind of language humans must be designed by evolution to produce and comprehend. It requires no special training or skill to be able to talk casually with others, and every normal person acquires this ability as a natural part of maturation. Because conversation is the form of language least influenced by acquired skills, it provides us with the most direct and uncontaminated access to natural mental processes." Chafe suggests that natural discourse should be of central importance to researchers inquiring into the cognitive

4

processes underlying language; I suggest that this is the case for inquiry into semantics as well. It might be argued that the types of mental processes and structures underlying semantics are fundamentally different from those underlying syntax, less innate and more cultural, and therefore call for a different type of data. However, the increasing amount of research which calls into question the possibility of a clear division between syntax and semantics (a position championed most prominently and persuasively by Goldberg (1998)) makes this position a challenging one to maintain.

Constructed examples and examples from written language *are* a valuable source of data in that they show the possible range of use of a word, and clearly no source of information should be ignored which sheds light on important and otherwise easily overlooked aspects of a words meaning. It isn't only important, however, to show the possible range of use of a word, although certainly having that knowledge is half the battle. Equally important is how people *do* use the word. Discourse data has the great advantage of showing not only the range of meaning that exists for a word, but the uses typical (or prototypical) of it. Knowledge of which syntactic frames, co-occurences with other words, and aspects of meaning are most typical of a word's use is central to understanding that word's semantics, and discourse analysis is the best source of such data.

Data from natural discourse, then, while certainly not the only useful source, should be of primary importance when looking at the meaning of lexical items.

4. SAY AND TELL

Say and *tell* are words which raise interesting issues because their semantic domains are different without being distinct. *Say* tends to be associated more with speech

itself, *tell* with the transfer of information underlying speech. Actually predicting, however, which word (or construction involving the word) is used in a sentence, based on this simple dichotomy, is difficult because of the large overlap in the domains of the two words. An approach for doing so is outlined in section 5; the data on which the approach is based (see appendix), consisting of three conversations, is taken from the Santa Barbara Corpus of Spoken American English.

Both *say* and *tell* are polysemous, and in both cases I'm dealing only with the verbal sense of the word. In the case of say, this omits the discourse particle *say* ("Say, why don't we go downtown") and the rhetorical use of *say* ("Say, for example, that you've lost all of your shoes. What do you wear on your feet?"). Only the speech act verb *say* ("I say you lost all your shoes on purpose") is included in the data. In the case of *tell*, this omits the directive *tell* ("He told me to take my shoes off") and the cognitive meaning of *tell* ("I can't tell if he's a madman or a genius…") and includes, again, only the speech-act verb meaning of tell ("…but I can tell you that he's not wearing any shoes").

4.1 Data

All of the data used in this study (see appendix for the complete data set) is taken from three conversations recorded in the Santa Barbara Corpus of Spoken American English (files 0trans_sbc1, 0trans_sbc2, and 0trans_sbc3). The data set below is the set of all sentences occurring in the three conversations which contain any verbal form of the words "say" and/or "tell."

Where syntactically coherent utterances are broken across speakers, I've connected the utterances and omitted the intervening dialogue.

6

[…] indicates an instance where a clear grammatical break from the preceding or following dialogue was difficult to identify, and for the sake of brevity only the relevant portion was included.

All examples used below are taken from the data set.

4.2 Tell

The most immediately apparent phenomenon in the data is that *tell* invariably occurs with both a subject and direct object•. This is consistent with Wierzbicka's definition of tell (1987: 286-287):

> I say: X
> I say this because I want to cause you to hear it.
> I assume this could cause you/someone to know it.

Wierzbicka's definition stresses that "telling" is an act in which the teller speaks with the intention that the person being told will hear, and thereby come to know something. *Tell* would seem to necessarily involve an agent (the teller) and a patient (the person being told), which explains well why *tell* would generally occur within an "X [tell] Y" syntactic frame.

Wierzbicka's definition asserts that *telling* is an act in which one speaker says something to another on the understanding that the person being told will hear, and thereby come to know, what is being told. This definition is consistent with all of the examples of *tell* in the data set, and her definition seems to be supported by discourse

---

• It may be that this is a shortcoming of the data set— the phrase, so popular among children, "I'm telling" might seem like an obvious counterexample. I would argue, however, that instances of *tell* occurring without a DO are generally limited to fixed phrases (well exemplified by "I'm telling") in which the comprising words have lost much of their individual meaning.

data. Particularly, Wierzbicka's analysis predicts the almost total absence of questions and exclamations in the data set:

> In any case, a component of this kind [I assume this could cause you/someone to know it] is necessary to account for the fact that not all kinds of utterances can be reported by means of *tell.* Declarative and imperative sentences usually can be so reported, but most questions and exclamations can't. For example, if X says to Y "What time is it?" or "How dare you?" these these utterances can hardly be reported by means of *tell,* even though the speaker clearly wants the addressee to hear him. The component "this could cause you/someone to know it" explains why *tell* cannot be used here (because what would *it* refer to?).

Of the 29 *tell* sentences in the data set, only 1 is not declarative, sentence 9, "Did mom tell you about Deb at all?" This question is allowed by Wierzbicka's definition because a question is not the thing being told. The sentence "Jeff told Mary 'did mom tell you about Deb at all'" would, presumably, not be allowed.

4.3 Say

No definition of *say* is offered in the NSM literature because SAY is included in the list of proposed semantic primitives, and is thus an "indefinable."

Wierzbicka (1972: 15) writes that "Indefinables are the bricks from which all human utterances are constructed, and as such they cannot belong to any kind of scientific or elitist jargon, but rather must be known to everyone, including children," and that "Indefinables should correspond to colloquial words (expressions) drawn from natural language."

It does not follow, however, from these two statements that the real-language counterparts of semantic primitives are themselves indefinable, because 1) for lexical primitives to be "known to everyone, including children," does not require that this

knowledge be crystallized in language, and 2) that semantic primitives have counterparts in real languages does not mean that those counterparts *are* the semantic primitives, only that they closely *correspond* to semantic primitives.

To define a lexical item "x" which corresponds to a semantic primitive X, it may only be necessary to say that "x" means X. *This is not the same as saying that "x" is undefinable.* "x" may be quite easy to define, but it remains in need of a definition. The semantic primitives upon which NSM definitions are based should indeed be seen as the building blocks of all lexical items, "the bricks from which all human utterances are constructed," but the words in human languages which most closely correspond to semantic primitives should not be confused with the semantic primitives themselves any more than the allophones of a phoneme should be confused with the phoneme itself. This conception of semantic primitives allows that they are universal, fundamental concepts to human cognition which are the building blocks if human language. The only criticism being offered here is that no items in the lexicons of real languages should be considered indefinable, because primitives should not be seen as directly manifested in human languages, but are rather as closely reflected in human languages.

Say is precisely such a lexical item: corresponding to, but not itself, a semantic primitive. The definition of the English word *say*, then, is SAY. A minor change, but not an ineffectual one: by bringing *say* into the realm of the definable, we place it on the same level as other members of the lexical set to which it belongs (i.e., *tell*) allowing a comparison of the two. It allows, moreover, for inquiry into the phenomena associated with its use in discourse (why, for example, does *say* rarely take a patient? Why does it introduce quotations so much more frequently than *tell* does?), which must be grounded

in its definition, and which thus could only otherwise be explained as "it just is"— an unsatisfying account at best.

4.4 Analysis

For a comparison of the two, an approach which looks at semantically relevant units is more useful than one which looks only at grammatical categories. "I said it" and "I told you it" seem like comparable sentences in terms of the information being presented, yet "it," in both cases referring to an utterance which has presumably previously been established as a referent in the conversation, has a different grammatical role in each of the two sentences.

For comparison of word choice within a lexical set, a schema that outlines the type of information being profiled is far more useful than one which outlines the grammatical role of the constituents. I'm operating on the assumption that words within a lexical set, because they perform similar functions in utterances, share a single schema. In the case of *tell* and *say*, I propose the schema:

<div align="center">Agent   V   Patient   U</div>

Agent refers to the person doing the telling or the saying, V to the verb being used (here, *tell* or *say*) patient to the person or thing being told something or the person or thing who something is being said to (the addressee), and U to the utterance which is being "told" or "said." U need not take the form of the actual utterance, but instead here stands for the phrase which references the utterance. In the sentence "People tell me stuff," 'people' is the agent, 'tell' is the V, 'me' the patient and 'stuff' the U. In "You said you got a book too" the agent is 'you', the V is "said," the patient is unspecified, and the U is 'you got a book too.'

Several forms of U's are attested in the data. U often takes the form of a quote, an attempt to recite verbatim an earlier utterance. Evidence that the speaker is quoting comes when the speaker adopts tense and person appropriate to the time and speaker of the original utterance ("We have our instructor right there, telling us, this is what you do,") or when the speaker uses a discourse marker (okay, well) to indicate that they are starting a quotation ("You know, we'd like, look at a horse's hoof, and say, okay, this is a double-aught.") Evidence that the speaker is paraphrasing comes when the speaker keeps the tense and person of the U appropriate to their own point of view rather than that of the original speaker— for example, "She said she really would love to see the turtles eat a goldfish" as opposed to "She said I really would love to see the turtle eat a goldfish." In many cases, no evidence is available in either direction and the U is ambiguous as to whether it's a quote or paraphrase.

In the performative use of *say* or *tell,* in which the speaker uses *say* or *tell* to signal that what they are saying or telling will follow ("Boy, I tell you, she's got arms the size of— they're huge"), U does not refer to or describe and earlier piece of discourse, but is itself the unit of discourse. Frequently, U takes the form of a pronoun which can have either a specified reference, referring to an utterance which occurred earlier in the discourse and has been established as a referent ("I don't know why I told you that"), or an unspecified reference, usually in the form of an indefinite pronoun ("Um, what was I gonna say?"). In some cases U is characterized without actually being paraphrased, the vital information contained (or which would be contained) in the utterance specified ("Just they held up the book but they didn't say who wrote it," "You said you wanted to

tell me about that Christmas card."). And finally, U, like agent and patient, is not always

specified ("I… I told you" and "He used to say to my brother, and he really meant it").

The frequencies in the data set of the different forms that U can take, as well as of

the different persons of the patient and agent, are listed in the table below. The

percentages of "U is a quote" and "U is a paraphrase" are presented as a range because of

the large number of sentences which offered no basis for judging whether the U of the

sentence in question was a quote or a paraphrase (17% for *tell*, 13% for *say*).

	Tell		Say	
	%	tokens	%	tokens
1'st person agent	55	16	34	20
2'nd person agent	14	4	17	10
3'rd person agent	31	9	49	29
No agent specified	0	0	0	0
1'st person patient	17	5	0	0
2'nd person patient	59	17	0	0
3'rd person patient	24	7	2	1
No patient specified	0	0	98	58
U is a quote	7-24	2-7	24-50	22-30
U is a paraphrase	3-21	1-6	5-31	11-19
U is a pronoun which refers to an utterance which has already been established as a referent in the discourse.	7	2	8	5
U characterizes the utterance	31	9	5	3
U has an unspecified referent	0	0	10	6
Performative use of *say*/*tell* introduces U	14	4	3	2
U does not occur	24	7	2	1
Total[*]	100	29	100	59

Tell and *say* seem to differ most prominently in the kind of information that is

being profiled (following Langacker, 1987). The greatest disparity between *tell* and *say*

[*] Totals represent subtotals (different types of agent, different types of patient, and different types ofU), not the sum of all figures in the category, which would yield an artificially high number. "Say" percentages are figured out of 59, not 57, because sentences 58 and 59 each represent two uses of "say."

in the table above is in the "no patient specified" category: *tell* invariably takes a patient; *say* occurs with a patient only once in the data set (sentence 58). Agent, on the other hand, is specified in all sentences for both *tell* and *say*. *Tell*, as predicted by Wierzbicka's definition, profiles both agent and patient, while *say* seems to profile only agent.

In terms of the types of agent specified, *tell* most frequently takes a 1'st person agent, followed (in descending order) by 3'rd person patient and 2'nd person patient, while *say* most frequently takes a third person agent, followed in frequency by 1'st and 2'nd person agents. As to patient, *tell* most frequently takes a 1'st person patient, 3'rd person second most frequently, and 1'st person least frequently. The one occurrence of *say* with a patient is with a 3'rd person patient.

Types of utterances were difficult to rank because of the large number of sentences in which there was no basis for making a judgement as to whether U represented a quote or a paraphrase. For both *tell* and *say* this makes it impossible to judge whether quotes or paraphrases are more common. For the sake of a firm ranking, both here and in the constraints set forth in section 5, "U is a quote" and "U is a paraphrase" are collapsed into a single category, "U is a quote or paraphrase."

The types of U's that occur with *tell* are, then, in descending order: U characterizes the utterance, U is a quote or paraphrase, U does not occur, U is the utterance itself, U is a pronoun, U has an unspecified referent.

The types of U's that occur with *say* are, in descending order: U is a quote or paraphrase, U has an unspecified referent, U is a pronoun, U characterizes the utterance, U is the utterance itself, and U does not occur.

5. AN OT ANALYSIS

5.1 Optimality Theory

Optimality Theory (Prince & Smolensky 1993), which over the decade since it was introduced has become a widely practiced approach to both phonology and syntax, is based on the principles that (Legendre 2001):

   a. UG is an optimizing system of universal well-formedness constraints on linguistic forms.
   b. Well-formedness constraints are simple and general. They routinely come into conflict and are (often) violated by the surfacing forms.
   c. Conflicts are resolved through hierarchical rankings of constraints. The effects of a given constraint is relative to its ranking, which is determined on a language-particular basis.
   d. Evaluation of candidates by the set of constraints is based on strict domination. For any two constraints C_1 and C_2, either C_1 outranks C_2 or C_2 outranks C_1.
   e. Alternative structural realizations of an input compete for the status of being the optimal output of a particular input. The most harmonic output— the one that best satisfies, or minimally violates, the full set of ranked constraints in a given language— is the optimal one. Only the optimal structure is grammatical.
   f. Every competition yields an optimal output.

In an OT analysis, it is assumed that the Generator (GEN) produces the set of all possible candidates, which are then run through the Evaluator (EVAL), the function of which is to judge which of the possible outputs is most "optimal"— the output which violates the least and/or lowest ranked of the constraints which constitute the EVAL. The model is presented in tableaux which present the candidates and constraints relevant to the issue at hand.

5.2 The application of OT to Semantics

The data presented above establishes that there is no absolute co-occurrence between the use of "say" and particular syntactic frame or the use of "tell" and a

particular syntactic frame in discourse, nor can NSM definitions alone provide a basis for understanding the means by which one is chosen over the other. A constraints and optimality model, providing as it does a model for understanding how many factors (constraints) can interact to decide the final output, seems ideally suited to an understanding of how one is chosen over the other in discourse.

The application of OT to semantics and lexical choice requires that each of the structural elements of OT (GEN, EVAL, constraints, candidates) have a corresponding element in the proposed model. In the model proposed here:

The candidates produced by GEN are the set of sentences in which each possible lexical choice for a particular position in the sentence is represented. The set would theoretically constitute a sentence for every member of a language's lexicon, making for an impractically large set. In practice, however, the vast majority of these candidates are immediately eliminated by extremely highly ranked (and thus never violated) constraints the function of which are to dictate that a particular position is filled by a member of a particular syntactic category (Noun, verb, etc.) and which dictate the semantic category which is suited to the intent of the speaker (these constraints being supplied by the elements of NSM definitions— see below). Pragmatically, only members of a particular lexical set need be considered.

Syntactic properties, moreover, of a word are taken into account, so that the candidate sentences are all sentences which would be acceptable to a native speaker— thus, for example, in the analysis below, the candidates considered are "People tell me stuff" and "People say stuff to me," not "People tell me stuff" and "People say me stuff."

The constraints which comprise the EVAL are the elements of NSM definitions and the patterns of a word's use, taken from discourse data. The full set of constraints would need to be immense, every element of every NSM definition and every pattern associated with every word's use accounted for. Again, however, only those relevant to the problem at hand need be considered, and thus only constraints relevant to the lexical set being dealt with are included in the analysis below.

5.3 *Tell* and *Say*

Say and *tell*, to return to the issue at hand, are words the use of which are difficult to predict based on their NSM definitions alone. The definitions of *say* and *tell* arrived at in section 4.2 are, respectively:

SAY

and

I say: X
I say this because I want to cause you to hear it.
I assume this could cause you/someone to know it.

The problem of predicting use based on these definitions alone is that the domains of the two definitions overlap, leaving a large area in which there is no basis for deciding which word is to be used. Why, for example, in sentence 42, is "She said people were driving around with their lights on" chosen over "She told me people were driving around with their lights on." The answer, of course, is that other factors are at work as well. Optimality Theory provides a model by which the interaction of the various factors can be understood, and discourse data a justification for the ranking of the constraints.

In ranking the constraints, the operative principle is that the phenomena connected with a word which occur *most* indicate the constraints which are violated *least*. Thus,

from the fact that *say* occurs without a patient 98% of the time in the data set, with a third person agent 49% of the time, and a U which is a pronoun 8% of the time, we can infer that "Say occurs without a patient" is a higher ranked constraint than "Say takes a third person agent, which in turn is ranked higher than "Say takes a U which is a pronoun."

NSM definitions are crafted with the intent of capturing the entire range of a word's use, covering every instance of its application. Thus every instance of "say" is covered by the definition SAY, and every instance of "tell" by its definition. On the principle above, if there is an absolute co-occurrence between particular instances of words and their definitions, then the constraints derived from the elements of definitions are extremely highly ranked, and thus never violated. The function of these constraints is to exclude "Colorless green ideas sleep furiously" sentences (allowed by constraints the function of which is to exclude ungrammatical sentences), by limiting on semantic grounds the possible set of words which can occur in a position. Thus we don't see the sentence "Bobby said the ball over the fence" because it violates the highly ranked constraint that "say" means SAY, a meaning incongruous with the intended semantics of the sentence. The function, then, of constraints derived from the elements of NSM definitions are to semantically limit the number of candidates to a small set, effectively to a lexical set such as "Verbs of speech."

Lowest in the hierarchy of constraints come those derived from the discourse data, which are violated to varying degrees in the data set. Semantic Tableaux, then, will take the general form:

Candidate sentences	Constraints which exclude ungrammatical sentences	Constraints derived from elements of NSM definitions	Constraints derived from discourse data
Sleep furiously ideas colorless green.	*!		
Colorless green ideas sleep furiously.		*!	

On this model, then, can the problem of choice within lexical sets, and thus the problem of how *say* is chosen over *tell* (or vice-versa) in discourse, can be approached. In sentence 17, for example, "People tell me stuff," the selection of *tell* over *say* can be understood in terms of the following tableau[*]:

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
☞	People tell me stuff						*	*						*	*	*			*	*			*				
	People say stuff to me					*!			*		*							*			*	*			*	*	*

Constraints:
1. "Tell" necessitates: I say: X
2. "Tell" necessitates: I say this because I want to cause you to hear it.
3. "Tell" necessitates: I assume this could cause you/someone to know it.
4. "Say" necessitates: I say: X.
5. Say does not take a patient
6. Tell takes a 2P patient
7. Tell takes a 1P agent
8. Say takes a U which is a quote or paraphrase
9. Say takes a 3P agent

---

[*] Constraints pertaining to grammaticality have been left out both for brevity and simplicity's sake, and on the assumption that none of them are violated in this scenario, and thus are not relevant to the analysis. Dotted lines represent borders between equally ranked constraints, derived from phenomena which occur with equal regularity in the data set. Elements of the NSM definition are equally ranked ranked because none of them are violated in the data set, providing no basis for ranking. The four criteria of the form "say/tell takes a U which is a quote/paraphrase" have been collapsed into the two criteria "say/tell takes a U which is a quote or paraphrase" in order to allow for a firm ranking of the constraints.

10. Say takes a 1P agent
11. Tell takes a U which characterizes the utterance.
12. Tell takes a 3P agent.
13. Tell takes a U which is a quote or paraphrase.
14. Tell does not take a U.
15. Tell takes a 3P patient.
16. Tell takes a 1P patient
17. Say takes a 2P agent.
18. Tell is used performatively, introducing U.
19. Tell takes a 2P agent.
20. Say takes a U which has an unspecified referent.
21. Say takes a U which is a pronoun that refers to a piece of discourse.
22. Tell takes a U which is a pronoun that refers to a piece of discourse.
23. Say takes a U which characterizes the utterance.
24. Say is used performatively, introducing U.
25. Say takes a 3P patient.
26. Say does not take a U.

     The constraints, as outlined above, are ranked in order of the frequency with which they are violated in the data set. "People say stuff to me" is eliminated by the fifth ranked criteria, "Say does not take a patient." As the highest-ranked criteria which is violated, it is the deciding factor in this instance. In sentence 79, "You said you got a book too," the fatal violation which dictates that *say* as chosen over *tell* comes a bit further down the line, at "Tell takes a 2P patient."

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
	You told me you got a book too.						*!	*				*	*		*	*	*		*				*				
☞	You said you got a book too.									*	*										*	*		*	*	*	*

     Sentence 58, "And he used to say to my brother, you know my brother was heavily into drug abuse, and… and he w— he— he used to say to my brother, and he really meant it," is a problematic one for the analysis because this use of *say* violates the

most highly ranked of the constraints derived from the data: Say does not take a patient. The problem is resolved by looking at the full context of the sentence:

> And he used to say to my brother, you know my brother was heavily into drug abuse, and .. and, he w- he— he used to say to my brother, and he really meant it. It sort- sort of like was empty words to teenagers. You know. But he really meant it, he said, you're real lucky, he said you've been blessed with a great body, and here you are fucking it up.

This is not an instance in which the semantic domains of *tell* and *say* overlap— *Say* is chosen over *tell* not on the basis of the constraints derived from discourse data, but on the basis of the constraints NSM definitions which outrank them. The use of *tell* would, in this instance, violate the criteria "Tell necessitates: I assume that this could cause you/someone to know it." Since the words being spoken are but "empty words to teenagers," the patient someone who will not come to know something as a result of the utterance, the speaker chooses "say" over "tell," and the violation of the lower ranked constraint "say does not take a patient" is irrelevant.

6. CONCLUSION

The model undoubtedly requires fine-tuning. Most importantly, the constraints undoubtedly could be less-word specific and more universal- for example, a constraint stating that verbs of cognition take a first person agent would be more universal than one which states that "think" takes a first person agent.

This study is a limited one; one which dealt with more data would offer further insight into how the criteria should be ranked and which criteria are most important. This study, moreover, looks at only a few variables, not necessarily the only important ones. Tense and aspect, for example, may have a role in the hierarchy of constraints. The

particular factors looked at and constraints chosen may leave much to be debated, but the validity of the approach— utilizing an OT system of constraints and optimality as model for lexical choice— holds.

Finally, an alternate approach to explicating the differences in distribution between the two words might be viewed in terms of the prototypical usage of each of the two words. For example, one might say, based on the data, that *tell* prototypically occurs with a 1'st person agent and a second person patient. An OT analysis does not contradict such a view, but rather should be seen as the model by which the various aspects of a word's profile (for example, the person of the agent and the type of utterance) are integrated into a single system by which the various semantic demands of an utterance are matched to the verb the prototypical use of which most closely matches them.

7. BIBLIOGRAPHY

CHAFE, W. (1998) Language and the Flow of Thought. In M. TOMASELLO (Ed.) *The new psychology of language.* Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

CUMMING, S. & ONO, T. (1997). Discourse and Grammar. In VAN DIJK, T. (Ed.) *Discourse Studies: a multidisciplinary introduction.* Thousand Oaks, CA: SAGE Publications.

GODDARD, C. (1998). *Semantic Analysis.* New York: Oxford University Press.

GODDARD, C. & WIERZBICKA, A. (1994). Introducing Lexical Primitives. In GODDARD, C. & WIERZBICKA, A. (Eds.) *Semantic and Lexical Universals: Theory and Empirical Findings.* Philadelphia, PA: John Benjamins Publishing Co.

GOLDBERG, A. (1998). Constructions: A Construction Grammar Approach to Argument Structure. In M. TOMASELLO (Ed.) *The new psychology of language.* Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

LANGACKER, R. W. (1987). *Foundations of cognitive grammar: Vol.1. Theoretical prerequisites.* Stanford, CA: Stanford University Press.

LEGENDRE, G. (2001). An Introduction to Optimality Theory in Syntax. In LEGENDRE, G., GRIMSHAW, J., & VIKNER, S. (Eds.) *Optimality-Theoretic Syntax.* Cambridge, MA: MIT Press.

PRINCE, A. & SMOLENSKY, P. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar.* Technical Report TR-2, Center for Cognitive Science, Rutgers University, New Brunswick, NJ., and Technical Report CU-CS-696-93, Department of Computer Science, University of Colorado, Boulder. Available at http://www.ruccs.rutgers.edu/publicationsreports.html.

WIERZBICKA, A. (1972). *Semantic Primitives.* Frankfurt/M.: Athenaum-Verl.

WIERZBICKA, A. (1987) *English speech act verbs: a semantic dictionary*. Sydney: Academic Press.

WIERZBICKA, A. (1998) The Semantics of English Causative Constructions in a Universal-Typological Perspective. In M. TOMASELLO (Ed.) *The new psychology of language.* Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Santa Barbara Corpus of Spoken American English (SBCSAE). Accessed April 3, 2003 at http://www.talkbank.org/data/conversation/SBCSAE/

8. APPENDIX: DATA

8.1 Tell
Conversation 1:

1. …you know, because he kinda has to tell you, they go over the whole, the whole part the horse.

2. …we have our instructor right there, telling us, this is what you do.

3. I was gonna tell you about that.

4. Boy I tell you, she's got arms the size of— they're huge.

5. And boy I tell you, you just— sweat just rolls off of you.

6. I was telling May, uh, day before yesterday, it was the most weird day I've ever seen in my life.

7. I… I told you.

8. I don't know why I told you that.

Conversation 2:

11. You were telling me about that.

12. I mean the guy is great… I told you.

13. I told you.

14. I told you.

15. I always told her, it's gonna, and she's⸺

16. And you must have been telling Mary or Dory or somebody.

17. People tell me stuff.

Conversation 3:

23. That the one I was telling you about, yeah.

24. I tell you I'm gonna have to be, really be….

25. I'd call the health department and tell them it's a health hazard.

26. Oh, why don't I let t- him tell you because it was something.

8.2 Say

Conversation 1:

30. I don't know how to say it.

31. Um, what was I gonna say?

9. Did mom tell you about Deb at all?

10. She's telling her mom, like, If I go I want this here, and I want that there, and you know, just little things.

18. I need to tell her first if she sees canister first she can do that too.

19. Hold on, Mary is going to tell you about her books.

20. Just a minute, I'm going to let Mary tell you.

21. I told Brittany I wish they'd given the address you can send and write off for it.

22. And then when they're done they don't tell you what it is again.

27. You said you wanted to tell me about that Christmas card.

28. You said you wanted to tell me about that Helen Stiener.

29. Well I was going to tell you about it but she beat me to it.

32. What was I gonna say…

33. You said you never made the horseshoes.

34. You know, we'd like, look at a horse's hoof and say, okay, this is a double-aught.

35. What was I gonna say?

36. I heard everybody saying, um, you know like the people that took the class before, talking about Oh, you have to do this, you have to do that…

37. You know, he said there was four guys doing that?

38. How he said there's four guys, I wonder if that means like, since there's four seasons, you know?

39. I got a kick out of him when he kept saying it.

40. I don't wanna say kitten, that sounds stupid.

41. And they were talking about the weather and they said how we—everybody woke up with ash all over their car today.

42. And they said well everybody woke up with ashes on their cars today.

43. She said people were driving around with their lights on.

44. And she said like the lights looked a greenish color.

45. So she was saying it didn't matter what she did.

46. But anyway, um, what was I gonna say.

Conversation 2:

47. They say Lebanon is actually—they— it just became illegal.

48. I think the early seventies, the Supreme Court, overturned travel restrictions, on the um, on the passport, and said that the government couldn't, really, permit you from going anywh—uh, prevent you from going anywhere.

49. I mean, well but I'm saying, he's the only existing live member of the Peace and Freedom Party left.

50. I've went to the doctor, and he said, he just did a blood test, and said, yeah, well, your blood's all shot, and you have the liver of a ninety-year-old.

51. He named like half a dozen viruses and said, ah, I could do some more tests…

52. And it said, uh, with pepsin.

53. And she said something, and she was walking out of the room, and I didn't catch it.

54. And I said excuse me, am I, um…

55. She said, don't take these in public.

56. But I'm saying, is like, you know, as much as he's abused his liver, and all other…

other things in his life, he's still healthy as an ox.

57. He plays hard, as they say, and yet he wants…

58. And he used to say to my brother, you know my brother was heavily into drug abuse, and… and, he w— he— he used to say to my brother, and he really meant it.

59. He said you've been real lucky, he said you've been blessed with a great body, and here you are fucking it up.

Conversation 3:

64. Brittany said call at one-thirty.

65. And we said yeah it worked and hung up.

66. And then I says well don't answer this time, let it ring a couple of times.

67. I heard you— I heard you say something— uh, what did you say?

68. Uh you said something to the effect, something is wrong here.

69. I think that's exactly what you said. Something is wrong here.

70. I said let me run get in the bathroom and maybe it's something with this phone.

71. Eh I says we're going to wait a half an hour {laugh} and let my thumb heal.

72. We literally oh I said we're going to wear out the buttons on the phone.

60. You said as they get older, they eat more vegetable matter.

61. When you say it that way, I don't think I could stand it, either.

62. She said she really would love to see the turtles eat a goldfish.

63. Now she's saying how it was the worst.

73. And they said don't call out there, tie up the lines, blah blah blah.

74. I said oh my gosh, because I thought hers are prettier.

75. That I thought she needed that she had said she needed and I would be a little surprised if she hadn't remembered me saying oh you need a breadbox.

76. I when I opened it up I said oh I use these all the time.

77. Stephanie says yeah she stole my idea because that is what Stephanie had given me last year.

78. That is what everybody says.

79. You said you got a book, too.

80. Just they held up the book but they didn't say who wrote it.

81. I think they introduced the guy, but they didn't say he was the one who wrote the book.

82. I said hey of you want to be fair about it spread them around the neighborhood.

83. Uh all I mean I hate to say it.

84. And he says the people that— all the people that have rented this house— said section eight did uh, the section eight people kept up good with it and checked on people.

85. You said you wanted to tell me about that Christmas card.

86. You said you wanted to tell me about that Helen Stiener.