Incremental Processing and Acceptability

Glyn Morrill*
Universitat Politècnica de Catalunya

We describe a left-to-right incremental procedure for the processing of Lambek categorial grammar by proof net construction. A simple metric of complexity, the profile in time of the number of unresolved valencies, correctly predicts a wide variety of performance phenomena including garden pathing, the unacceptability of center embedding, preference for lower attachment, left-to-right quantifier scope preference, and heavy noun phrase shift.

1. Introduction

Contemporary linguistics rests on abstractions and idealizations which, however fruitful, should eventually be integrated with human computational performance in language use. In this paper we consider the case of language processing on the basis of Lambek categorial grammar (Lambek 1958). We argue that an incremental procedure of proof net construction affords an account of various processing phenomena, including garden pathing, the unacceptability of center embedding, preference for lower attachment, left-to-right quantifier scope preference, and heavy noun phrase shift. We give examples of each of these phenomena below.

Garden pathing (Bever 1970) is illustrated by the following contrasts:

(1) a. The horse raced past the barn.

b. ?The horse raced past the barn fell.

(2) a. The boat floated down the river.

b. ?The boat floated down the river sank.

(3) a. The dog that knew the cat disappeared.

b. ?The dog that knew the cat disappeared was rescued.

Typically, although the b sentences are perfectly well formed they are perceived as being ungrammatical due to a strong tendency to interpret their initial segments as in the a sentences.

The unacceptability of centre embedding is illustrated by the fact that while the nested subject relativizations of (4) exhibit little variation in acceptability, the nested

* Departament de Llenguatges i Sistemes Informatics, Modul C 5 - Campus Nord, Jordi Girona Salgado 1–3, E-08034 Barcelona. E-mail: morrill@lsi.upc.es; http://www-lsi.upc.es/~morrill/

object relativizations (5) exhibit a severe deterioration in acceptability (Chomsky 1965, Chap. 1).

(4) a. The dog that chased the cat barked.

b. The dog that chased the cat that saw the rat barked.

c. The dog that chased the cat that saw the rat that ate the cheese barked.

(5) a. The cheese that the rat ate stank.

b. ?The cheese that the rat that the cat saw ate stank.

c. ??The cheese that the rat that the cat that the dog chased saw ate stank.

Discussing such center embedding, Johnson (1998) presents the essential idea developed here, noting that processing overload of dependencies invoked in psycholinguistic literature could be rendered in terms of the maximal number of unresolved dependencies as represented by proof nets.

Kimball (1973, 27) considers sentences such as (6), which are three ways ambiguous according to the attachment of the adverb. He points out that the lower the attachment of the adverb, the higher the preference (he calls this relationship Right Association).

(6) Joe said that Martha believed that Ingrid fell today.

Left-to-right quantifier scope preference is illustrated by:

(7) a. Someone loves everyone.

b. Everyone is loved by someone.

Both sentences exhibit both quantifier scopings:

(8) a. $\exists x \forall y (love\ y\ x)$

b. $\forall y \exists x (love\ y\ x)$

However, while the dominant reading of (7a) is (8a), that of (7b) is (8b), i.e., the preference is for the first quantifier to have wider scope. Note that the same effect is observed when the quantifiers are swapped:

(9) a. Everyone loves someone.

b. Someone is loved by everyone.

While both sentences in (9) have both quantifier scopings, the preferred readings give the first quantifier wide scope.

Finally, we will look at heavy noun phrase shift, which is the preference for complex object noun phrases to "shift" to the end of the sentence. Consider the two sentences in (10); the second, in which the "heavy" direct object follows the indirect object, is more acceptable than the first.

(10) a. ?John gave the painting that Mary hated to Bill.

b. John gave Bill the painting that Mary hated.

We argue that a simple metric of categorial processing complexity explains these and other performance phenomena.

2. Lambek Calculus

We shall assume some familiarity with Lambek categorial grammar as presented in, for example, Moortgat (1988, 1997), Morrill (1994), or Carpenter (1998), and limit ourselves here to reviewing some central technical and computational aspects.

The types, or (category) formulas, of Lambek calculus are freely generated from a set of primitives by the binary infix connectives "/" (over), "\" (under) (directional divisions) and "·" (product). With respect to a semigroup algebra $(L, +)$ (i.e., a set L closed under an associative binary operation + of adjunction), each formula A is interpreted as a subset $[\![A]\!]$ of L by residuation as follows:

(11)
$$[\![A{\cdot}B]\!] = \{s_1{+}s_2 | s_1 \in [\![A]\!] \ \&\ s_2 \in [\![B]\!]\}$$
$$[\![A\backslash B]\!] = \{s | \forall s' \in [\![A]\!], s'{+}s \in [\![B]\!]\}$$
$$[\![B/A]\!] = \{s | \forall s' \in [\![A]\!], s{+}s' \in [\![B]\!]\}$$

A sequent, $\Gamma \Rightarrow A$, comprises a succedent formula A and an antecedent configuration Γ, which is a a finite sequence of formulas.[1] A sequent is valid if and only if in all interpretations the ordered adjunction of elements inhabiting the antecedent formulas always yields an element inhabiting the succedent formula. The following Gentzen-style sequent presentation is sound and complete for this interpretation (Buszkowski 1986, Došen 1992), and indeed for free semigroups (Pentus 1994): hence the Lambek calculus can make an impressive claim to be *the* logic of concatenation; a parenthetical notation $\Delta(\Gamma)$ represents a configuration containing a distinguished subconfiguration Γ.

(12)

a. $$A \Rightarrow A \quad id \qquad \frac{\Gamma \Rightarrow A \quad \Delta(A) \Rightarrow B}{\Delta(\Gamma) \Rightarrow B}\,\mathrm{Cut}$$

b. $$\frac{\Gamma \Rightarrow A \quad \Delta(B) \Rightarrow C}{\Delta(\Gamma, A\backslash B) \Rightarrow C}\,\backslash\mathrm{L} \qquad \frac{A, \Gamma \Rightarrow B}{\Gamma \Rightarrow A\backslash B}\,\backslash\mathrm{R}$$

c. $$\frac{\Gamma \Rightarrow A \quad \Delta(B) \Rightarrow C}{\Delta(B/A, \Gamma) \Rightarrow C}\,/\mathrm{L} \qquad \frac{\Gamma, A \Rightarrow B}{\Gamma \Rightarrow B/A}\,/\mathrm{R}$$

d. $$\frac{\Gamma, A, B, \Delta \Rightarrow C}{\Gamma, A{\cdot}B, \Delta \Rightarrow C}\,{\cdot}\mathrm{L} \qquad \frac{\Gamma \Rightarrow A \quad \Delta \Rightarrow B}{\Gamma, \Delta \Rightarrow A{\cdot}B}\,{\cdot}\mathrm{R}$$

By way of example, "lifting" A ⇒ B/(A\B) is generated as follows:

(13)
$$\dfrac{\dfrac{A \Rightarrow A \quad B \Rightarrow B}{A, A\backslash B \Rightarrow B}\,\backslash\mathrm{L}}{A \Rightarrow B/(A\backslash B)}\,/\mathrm{R}$$

1 Officially, the antecedent is nonempty, a detail we gloss over.

And "composition" A\B, B\C $\Rightarrow$ A\C can be derived thus:

(14)
$$\dfrac{\dfrac{A \Rightarrow A \quad \dfrac{B \Rightarrow B \quad C \Rightarrow C}{B, B\backslash C \Rightarrow C}\,\backslash\mathrm{L}}{A, A\backslash B, B\backslash C \Rightarrow C}\,\backslash\mathrm{L}}{A\backslash B, B\backslash C \Rightarrow A\backslash C}\,\backslash\mathrm{R}$$

Every rule, with the exception of Cut, where the Cut formula A does not appear in the conclusion, has exactly one connective occurrence less in its premisses than in its conclusion. Lambek (1958) proved Cut elimination—that every proof has a Cut-free counterpart—hence a decision procedure for theoremhood is given by backward-chaining proof search in the Cut-free fragment. The nonatomic instances of the id axiom are derivable from atomic instances by the rules for the connectives. But even in the Cut-free atomic-id calculus there is spurious ambiguity: equivalent derivations differing only in irrelevant rule ordering. For example, composition as above has the following alternative derivation:

(15)
$$\dfrac{\dfrac{\dfrac{A \Rightarrow A \quad B \Rightarrow B}{A, A\backslash B \Rightarrow B}\,\backslash\mathrm{L} \quad C \Rightarrow C}{A, A\backslash B, B\backslash C \Rightarrow C}\,\backslash\mathrm{L}}{A\backslash B, B\backslash C \Rightarrow A\backslash C}\,\backslash\mathrm{R}$$

One approach to this problem consists in defining, within the Cut-free atomic-id space, normal form derivations in which the succession of rule application is regulated (König 1989, Hepple 1990, Hendriks 1993). Each sequent has a distinguished category formula (underlined) on which rule applications are keyed:

(16) a. $\underline{P} \Rightarrow P \qquad id$

$$\dfrac{\Delta(\underline{A}) \Rightarrow B}{\Delta(A) \Rightarrow \underline{B}}$$

b. $$\dfrac{\Gamma \Rightarrow \underline{A} \quad \Delta(\underline{B}) \Rightarrow C}{\Delta(\Gamma, \underline{A\backslash B}) \Rightarrow C}\,\backslash\mathrm{L} \qquad \dfrac{A, \Gamma \Rightarrow \underline{B}}{\Gamma \Rightarrow \underline{A\backslash B}}\,\backslash\mathrm{R}$$

c. $$\dfrac{\Gamma \Rightarrow \underline{A} \quad \Delta(\underline{B}) \Rightarrow C}{\Delta(\underline{B/A}, \Gamma) \Rightarrow C}\,/\mathrm{L} \qquad \dfrac{\Gamma, A \Rightarrow \underline{B}}{\Gamma \Rightarrow \underline{B/A}}\,/\mathrm{R}$$

d. $$\dfrac{\Delta \Rightarrow \underline{A} \quad \Gamma \Rightarrow \underline{B}}{\Delta, \Gamma \Rightarrow \underline{A{\cdot}B}}\,\cdot R$$

In the regulated calculus there is no spurious ambiguity, and provided there is no explicit or implicit antecedent product, i.e., provided ·L is not needed, $\Gamma \Rightarrow A$ is a theorem of the Lambek calculus iff $\Gamma \Rightarrow \underline{A}$ is a theorem of the regulated calculus. However, apart from the issue regarding ·L, there is a general cause for dissatisfaction with this approach: it assumes the initial presence of the entire sequent to be proved, i.e., it is in principle nonincremental; on the other hand, allowing incrementality on the basis of Cut would reinstate with a vengeance the problem of spurious ambiguity, for then what are to be the Cut formulas? Consequently, the sequent approach is ill-equipped to address the basic asymmetry of language—the asymmetry of its processing in time—and has never been forwarded in a model of the kind of processing phenomena cited in the introduction.

An alternative formulation (Ades and Steedman 1982, Steedman 1997), which from its inception has emphasized a capacity to produce left-branching, and therefore incrementally processable, analyses, is comprised of combinatory schemata such as the following (together with a Cut rule, feeding one rule application into another):

(17) a. $A, A\backslash B \Rightarrow B \qquad B/A, A \Rightarrow B$

b. $A \Rightarrow (B/A)\backslash B \qquad A \Rightarrow B/(A\backslash B)$

c. $A\backslash B, B\backslash C \Rightarrow A\backslash C \qquad C/B, B/A \Rightarrow C/A$

By a result of Zielonka (1981), the Lambek calculus is not axiomatizable by any finite set of combinatory schemata, so no such combinatory presentation can constitute *the* logic of concatenation in the sense of Lambek calculus. Combinatory categorial grammar does not concern itself with the capture of all (or only) the concatenatively valid combinatory schemata, but rather with incrementality, for example, on a shift-reduce design. An approach (also based on regulation of the succession of rule application) to the associated problem of spurious ambiguity is given in Hepple and Morrill (1989) but again, to our knowledge, there is no predictive relation between incremental combinatory processing and the kind of processing phenomena cited in the introduction.

3. Proof Nets

Lambek categorial derivations are often presented in the style of natural deduction or sequent calculus. Here we are concerned with categorial proof nets (Roorda 1991) as the fundamental structures of proof in categorial logic, in the same sense that linear proof nets were originally introduced by Girard (1987) as the fundamental structures of proof in linear logic. (Cut-free) proof nets exhibit no spurious ambiguity and play the role in categorial grammar that parse trees play in phrase structure grammar.

Surveys and articles on the topic include Lamarche and Retoré (1996), de Groote and Retoré (1996), and Morrill (1999). Still, at the risk of proceeding at a slightly slower pace, we aim nonetheless to include here enough details to make the present paper self-contained.

A **polar category formula** is a Lambek categorial type labeled with **input** ($^\bullet$) or **output** ($^\circ$) polarity. A **polar category formula tree** is a binary ordered tree in which the leaves are labeled with polar atoms (**literals**) and each local tree is one of the following (**logical**) **links**:

(18) a. $\dfrac{A^\circ \quad B^\bullet}{A\backslash B^\bullet}\,\mathrm{ii} \qquad \dfrac{B^\circ \quad A^\bullet}{A\backslash B^\circ}\,\mathrm{i}$

b. $\dfrac{B^\bullet \quad A^\circ}{B/A^\bullet}\,\mathrm{ii} \qquad \dfrac{A^\bullet \quad B^\circ}{B/A^\circ}\,\mathrm{i}$

c. $\dfrac{A^\bullet \quad B^\bullet}{A{\cdot}B^\bullet}\,\mathrm{i} \qquad \dfrac{B^\circ \quad A^\circ}{A{\cdot}B^\circ}\,\mathrm{ii}$

Without polarities, a formula tree is a kind of formation tree of the formula at its root:

daughters are labeled with the immediate subformulas of their mothers. The polarities indicate sequent sidedness, input for antecedent and output for succedent; the polarity propagation follows the sidedness of subformulas in the sequent rules: in the antecedent (input) rule for $A\backslash B$ the subformula A goes in a succedent (output) and the subformula B goes in an antecedent (input); in the succedent (output) rule for $A\backslash B$ the subformula A goes in an antecedent (input) and the subformula B goes in a succedent (output); etc. The labels i and ii indicate whether the corresponding sequent rule is unary or binary. Note that in the output links the order of the subformulas is switched; this corresponds to a cyclic reading of sequents: the succedent type is adjacent to the first antecedent type.

A **proof frame** is a finite sequence of polar category formula trees, exactly one of which has a root of output polarity (corresponding to the unique succedent of sequents).

An **axiom linking** on a set of literal labeled leaves is a partitioning of the set into pairs of complementary leaves that is planar in its ordering, i.e., there are no two pairs $\{L_1, L_3\}, \{L_2, L_4\}$ such that $L_1 < L_2 < L_3 < L_4$. Geometrically, planarity means that where the leaves are ordered on a line, paired leaves can be connected in the half plane without crossing. Axiom links correspond to id instances in a sequent proof.

A **proof structure** is a proof frame together with an axiom linking on its leaves. A proof net is a proof structure in which every elementary (i.e., visiting vertices at most once) cycle crosses the edges of some i-link.[2] Geometrically, an elementary cycle is the perimeter of a face or cluster of faces in a planar proof structure. There is a proof net with roots $A^\circ, A_1{}^\bullet, \ldots, A_n{}^\bullet$ iff $A_1, \ldots, A_n \Rightarrow A$ is a valid sequent.

4. Incremental Processing Load and Acceptability

Let us assume the following lexical assignments:

(19)

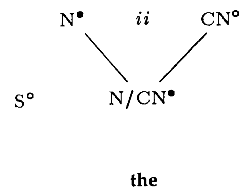
barn	–	*barn*
	:=	CN
horse	–	*horse*
	:=	CN
past	–	$\lambda x \lambda y \lambda z (past\ x\ (y\ z))$
	:=	$((\mathrm{N}\backslash\mathrm{S}t)\backslash(\mathrm{N}\backslash\mathrm{S}t))/\mathrm{N}$
raced	–	*race*
	:=	$\mathrm{N}\backslash\mathrm{S}+$
the	–	*the*
	:=	N/CN

The feature $+$ on S marks the projection of a tensed verb form; a verb phrase modified by *past* need not be tensed. Let us consider the incremental processing of (1a) as proof

2 This criterion, adapted from that of Lecomte and Retoré (1995), derives from Girard's (1987) long trip condition, which is an involved mathematical result. Danos and Regnier (1989) express it in terms of acyclicity and connectivity of certain subgraphs. Intuitively, acyclicity assures that the subformulas of ii-links (binary rules) occur in *different* subproofs, whereas connectivity assures that the subformulas of i-links (unary rules) occur in the *same* subproofs (attributed to Jean Gallier by Philippe de Groote, p.c.). However the single-succedent (intuitionistic) nature of Cut-free categorial proofs in fact renders the connectivity requirement redundant, hence we have just an acyclicity test.

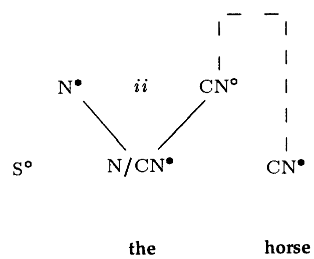
net construction.[3] In the first case, we suppose that one initially *expects* some target category, perhaps (though not necessarily) S. This ‘principle of expectation’ seems a reasonable or obvious principle of communication; as we shall see, it turns out to be technically critical. After perception of the word *the* there is the following partial proof net (for simplicity we omit features, included in lexical entries, from proof nets themselves):

(20)



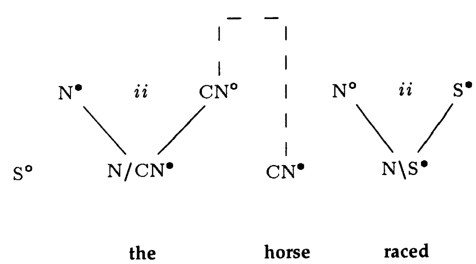
Here there are three unmatched valencies/unresolved dependencies; no axiom links can yet be placed, but after *horse* we can build:

(21)



Now there are only two unmatched valencies. After *raced* we have, on the correct analysis, the following:

(22)

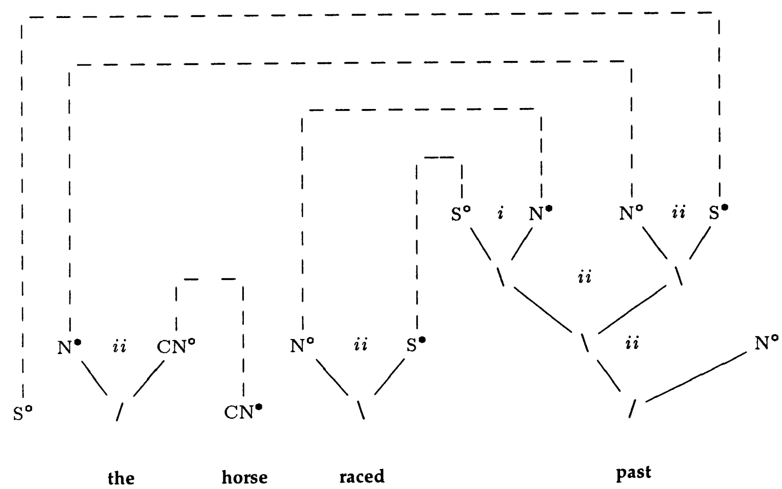


Note that linking the Ns is possible, but we are interested in the history of the *correct* analysis, and in that, the verb valencies are matched by the adverb that

3 The procedure is similar in spirit to that in the appendix of Ades and Steedman (1982), but we perform “reductions” by axiom links on complementary literal pairs rather than by combinatory schemata on category formulas.

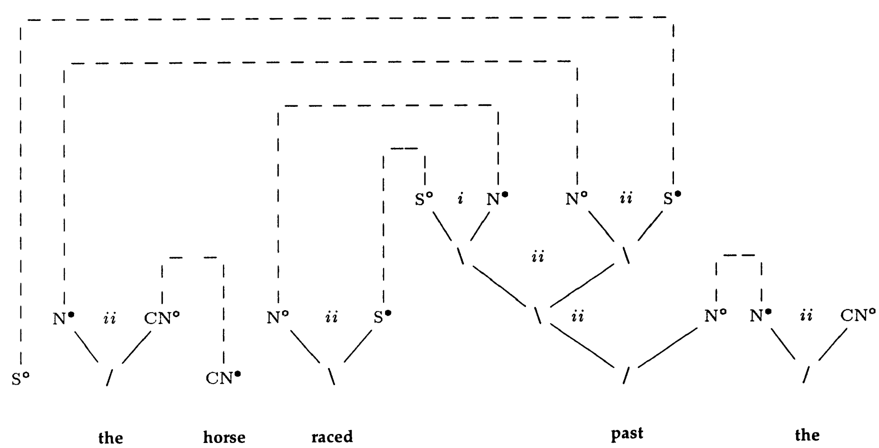
follows (henceforth we indicate only the principal connective of a mother node):

(23)



Observe that a cycle is created, but as required it crosses the edges of an i-link. At the penultimate step we have:

(24)



The final proof net analysis is given in Figure 1.

The semantics associated with a categorial proof net, i.e., the proof as a lambda term (intuitionistic natural deduction proof, under the Curry-Howard correspondence) is extracted by associating a distinct index with each output division node and traveling as follows, starting by going up at the unique output root (de Groote and Retoré 1996):

(25)
- traveling up at the mother of an output division link, perform the lambda abstraction with respect to the associated index of the result of traveling up at the daughter of output polarity;
- traveling up at the mother of an output product link, form the ordered pair of the result of traveling up at the right daughter (first component) and the left daughter (second component);

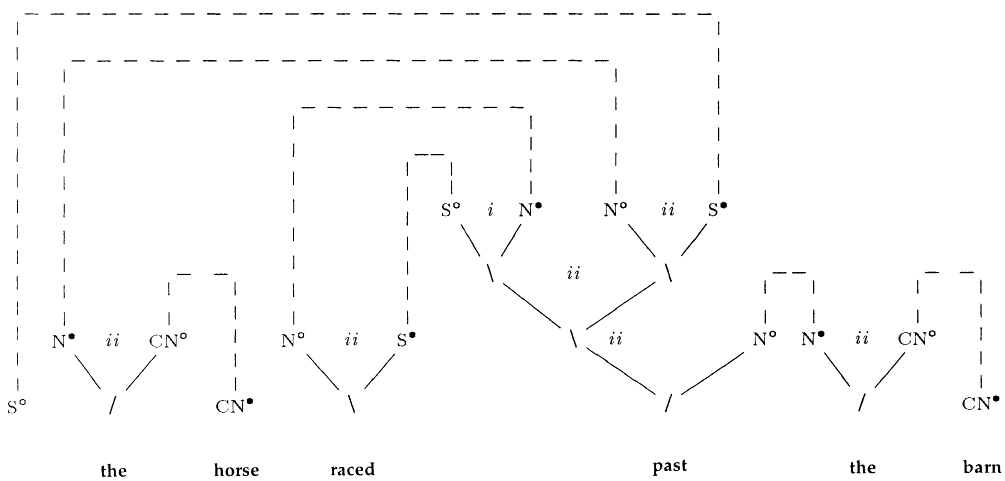


Figure 1
Proof net analysis of (1a) *the horse raced past the barn*.

- traveling up at one end of an axiom link, continue down at the other end;
- traveling down at an (input) daughter of an input division link, perform the functional application of the result of traveling down at the mother to the result of traveling up at the other (output) daughter;
- traveling down at the left (respectively, right) daughter of an input product link, take the first (respectively, second) projection of the result of traveling down at the mother;
- traveling down at the (input) daughter of an output division link, return the associated index;
- traveling down at a root, return the associated lexical semantics.

Thus for our example we obtain (26a), which is logically equivalent to (26b).

(26) a. $(\lambda x \lambda y \lambda z(\mathit{past}\ x\ (y\ z))\ (\mathit{the\ barn})\ \lambda 1(\mathit{race}\ 1)\ (\mathit{the\ horse}))$

b. $(\mathit{past}\ (\mathit{the\ barn})\ (\mathit{race}\ (\mathit{the\ horse})))$

The analysis of (1b) is less straightforward. Whereas in (1a) *raced* expresses a one-place predication ("go quickly"), in (1b) it expresses a two-place predication (there was some agent racing the horse); *horse* is modified by an agentless passive participle, but the adverbial *past the barn* is modifying *race*. Within the confines of the Lambek calculus, the characterization we offer assumes the lexical assignment to the passive participle given in the following:[4]

4 In general, grammar requires the expressivity of more powerful categorial logics than just Lambek calculus; however, so far as we are aware, the characterizations we offer within the Lambek calculus bear the same properties with regard to our processing considerations as their more sophisticated categorial logic refinements, because the latter are principally concerned with generalizations of word order, whereas the semantic dependencies on which our complexity metric depends remain the same.

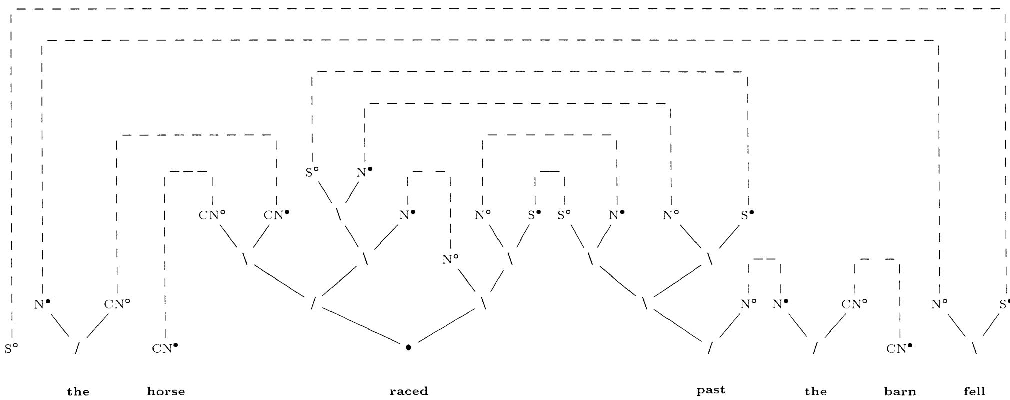


Figure 2
Proof net analysis of (1b) *the horse raced past the barn fell*.

(27) **fell** – *fall*
:= $N\backslash S+$
raced – $(\lambda x \lambda y \lambda z[(y\ z) \wedge \exists w(x\ z\ w)], race2)$
:= $((CN\backslash CN)/(N\backslash(N\backslash S-)))\cdot(N\backslash(N\backslash S-))$

Here *raced* is classified as the product of an untensed transitive verbal type, which can be modified by the adverbial *past the barn* by composition, and an adnominalizer of this transitive verbal type. According to this, (1b) has the proof net analysis given in Figure 2. The semantics extracted is (28a), equivalent to (28b)

(28) a. $(fall\ (the\ (\pi_1(\lambda x \lambda y \lambda z[(y\ z) \wedge \exists w(x\ z\ w)], race2)\ \lambda 29 \lambda 30$
$(\lambda u \lambda v \lambda w(past\ u\ (v\ w))\ (the\ barn)\ \lambda 41(($
$(\pi_2(\lambda p \lambda s \lambda t[(s\ t) \wedge \exists q(p\ t\ q)], race2)\ 29\ 41)\ 30)\ horse)))$

b. $(fall\ (the\ \lambda 8[(horse\ 8) \wedge \exists 7(past\ (the\ barn)\ (race2\ 8\ 7))]))$

Let us assign to each proof net analysis a complexity profile that indicates, before and after each word, the number of unmatched literals, i.e., unresolved valencies or dependencies, under process at that point. This is a measure of the course of memory load in optimal incremental processing. We are not concerned here with resolution of lexical ambiguity or serial backtracking: we are supposing sufficient resources that the nondeterminism of selection of lexical entries and their parallel consideration is not the critical burden. Rather, the question is: which among parallel competing analyses places the least load on memory?

Since entropy degrades the structure of memory, it requires more energy to pursue an analysis that is high cost in memory than to pursue one that is low cost. From these simple economic considerations we derive our main claim:

(29) **Principle of Acceptability**

Acceptability is inversely proportional to the sum in time of the memory load of unresolved valencies.

If other factors are constant, the principle makes a quantitative prediction. We can distinguish two cases: synonymy and ambiguity. In the case of synonymy, semantics is constant. It is then predicted that amongst synonymous forms of expression, the lower

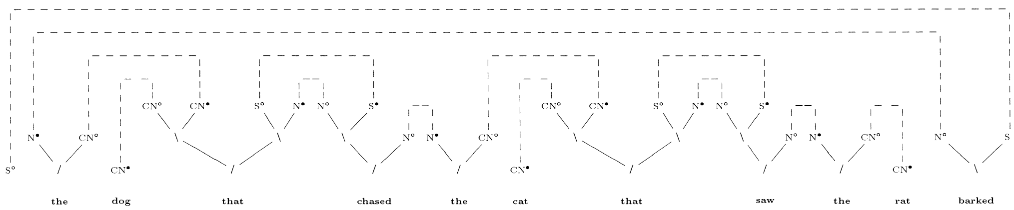


Figure 3
Proof net analysis of (4b) *the dog that chased the cat that saw the rat barked*.

the complexity curve, the higher the preference for the form of expression. In the case of ambiguity, prosodics is constant. It is then predicted that amongst the readings of an ambiguous expression, the lower the complexity curve, the more dominant the reading.

The complexity profile is easily read off a completed proof net: the complexity between two words is the number of axiom links bridging rightwards (forwards in time) at that point. Thus for (1a) and (1b) analyzed in Figures 1 and 2, the complexity profiles are as follows:

(30)

6				b				
5								
4			b	a				
3		ab			b	b		
2			a				b	
1	ab				a	a		
0							a	b
a.	the	horse	raced	past	the	barn		
b.	the	horse	raced	past	the	barn	fell	

We see that after the first two words the complexity of the locally ambiguous initial segment of (1b) is consistently higher than that of its garden path (1a). The areas of the a and b curves are 12 and 22 respectively, predicting that in (1b) the less costly but incorrect analysis could be salient, as indeed it is.

Johnson (1998) considers center embedding for subject and object relativization from a similar point of view. We assume here the relative pronoun lexical assignments shown in (31).[5]

(31)
$$\begin{array}{rcl} \textbf{that} & - & \lambda x \lambda y \lambda z[(y\ z) \wedge (x\ z)] \\ & := & (CN\backslash CN)/(N\backslash S+) \\ \textbf{that} & - & \lambda x \lambda y \lambda z[(y\ z) \wedge (x\ z)] \\ & := & (CN\backslash CN)/(S+/N) \end{array}$$

The proof net analysis of sentence (4b) is shown in Figure 3, and that of sentence (5b) is shown in Figure 4. Let us compare the complexities:

(32)

9								b				
8									b			
7						b	b					
6					b							
5										b		
4			ab	ab			a	a				
3		ab			a	a			a	a		
2											ab	
1	ab											
0												ab
a.	the	dog	that	chased	the	cat	that	saw	the	rat	barked	
b.	the	cheese	that	the	rat	that	the	cat	saw	ate	stank	

5 For better linguistic treatment not affecting the point at hand, see Morrill (1994, Chap. 8).

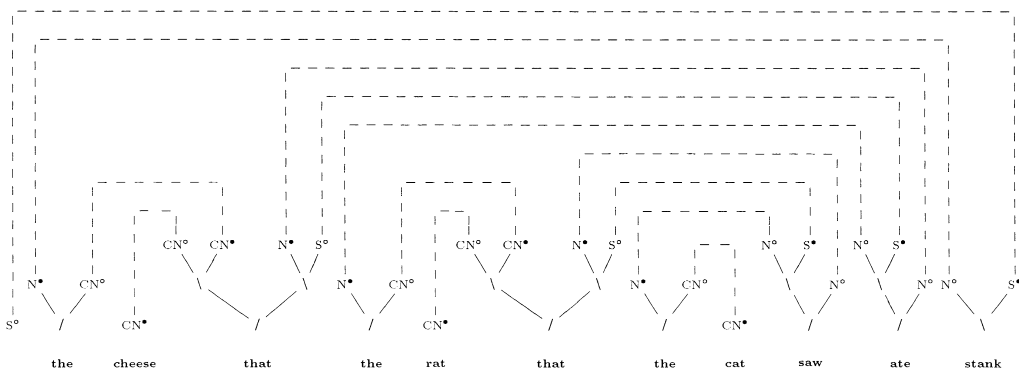


Figure 4
Proof net analysis of (5b) *the cheese that the rat that the cat saw ate stank*.

The profile of (5b) is higher; indeed it rises above 8, thus reaching what is usually taken to be the limit of short-term memory. Johnson attributes the increasing ill-formedness of centre embedded constructions to the number of incomplete dependencies at the "maximal cut" of a proof net. This almost corresponds to the maximum height of a complexity profile here, except Johnson includes no target category, whereas we will argue in relation to quantifier scope preference that this is critical. However, we also differ from Johnson in attributing relative acceptability to the area under the complexity curve, not only its maximal height. This is because we believe acceptability is to be explained in terms of the energy required to maintain processes in memory over time, and not just in terms of peak memory load. Finally, it happens that our proposal solves a problem encountered by Johnson. Gibson and Thomas (1996) observe that (33a) is easier to comprehend than (33b).

(33) a. The chance that the nurse who the doctor supervised lost the reports bothered the intern.

b. ?The intern who the chance that the doctor lost the reports bothered supervised the nurse.

Johnson notes that his proposal does not capture this difference since both sentences have the same size maximal cut. Under our account, on the other hand, it is the complexity curves as a whole that account for acceptability. In these sentences, although the height is the same, the complexity curves are not: the area of (33a) is less than that of (33b).[6] Thus, whereas Johnson must look to other factors to explain this difference, our account makes the correct prediction.

6 To save space we exclude the proof nets; the curves are:

8								ab								
7									ab							
6					b	ab	ab			b	b					
5					a							b				
4			b	b						a						
3		ab	a	a							a	a				
2													ab			
1	ab													ab	ab	
0																ab

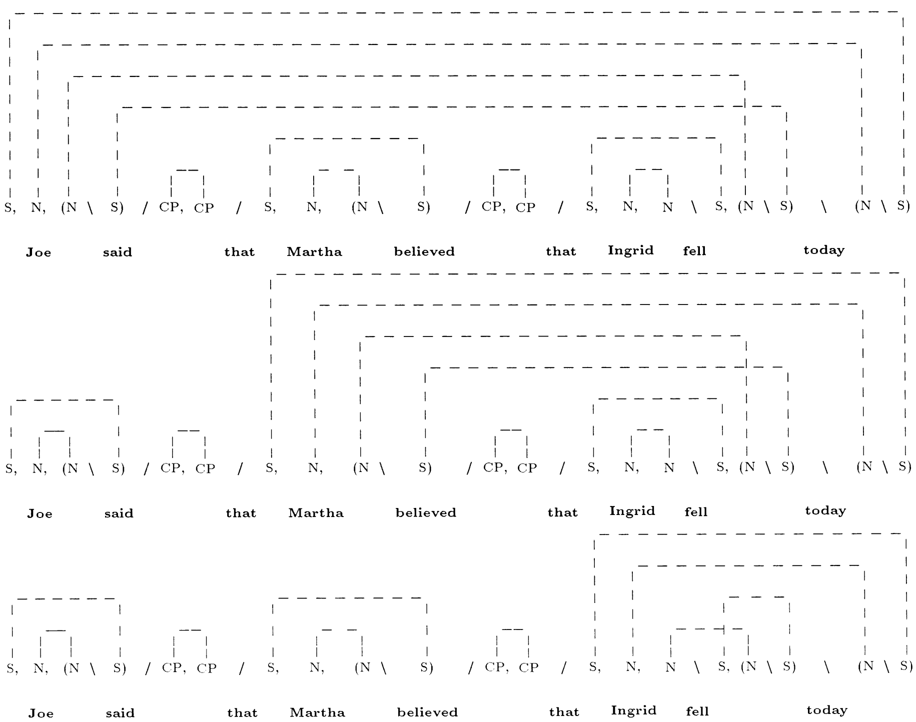


Figure 5
Proof net analyses for (34) *Joe said that Martha believed that Ingrid fell today*, with lowest (top), middle (center), and highest (bottom) attachment of the adverb *today*.

5. Preferred Readings

The examples so far involve comparison of different expressions, having many differences to which their relative acceptabilities could be attributed. More factors would be held constant by comparing alternative readings of an ambiguous expression, and most appropriately of all, a structurally ambiguous expression, where there is no lexical alternation. Consider the ambiguity that arises in a sentence such as (34) (repeated from (6)) from the possibility of attaching the adverb at different syntactic levels. As we saw in the introduction, the lower the attachment of the adverb in such sentences, the higher the preference (Kimball 1973, 27).

(34) Joe said that Martha believed that Ingrid fell today.

In Figure 5 we give the analyses for the lowest, the middle, and the highest attachments. We now abbreviate proof nets by flattening formula trees into their linear representations (since this conceals the order switching of output links, the notation belies the underlying planarity, but the portrayal of word-by-word complexity is unaltered). Accordingly, the complexity profiles are:

(35)

6					c			bc		
5			c	c		bc	bc			
4									abc	
3										
2		abc			ab			a		
1	abc		ab	ab		a	a			
0										abc
a.										lowest attachment
b.	Joe	said	that	Martha	believed	that	Ingrid	fell	today	middle attachment
c.										highest attachment

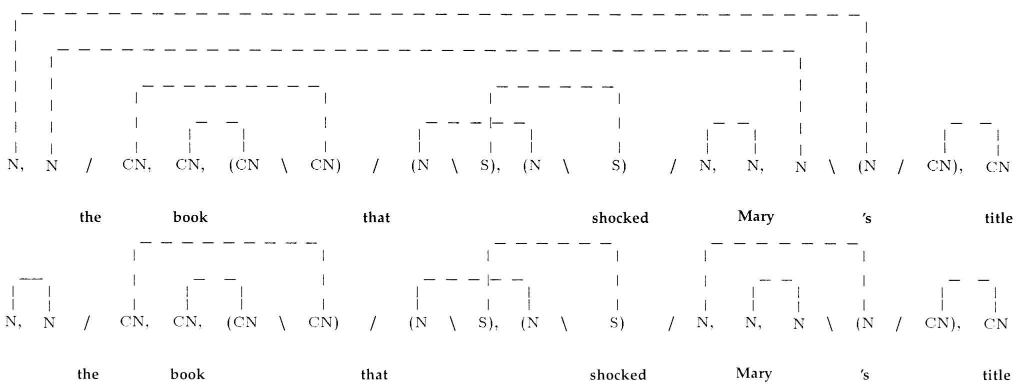


Figure 6
Proof net analysis for the sensical (top) and nonsensical (bottom) interpretations of (36) *the book that shocked Mary's title*.

The same effect occurs strongly in (36), where the preferred reading is the one given by the lowest attachment, even though that one is the nonsensical reading.

(36) the book that shocked Mary's title

The analyses are given in Figure 6. The complexities are thus:

(37)

4			a	a				
3		a			a			
2			b	b		ab		
1	ab	b			b		ab	
0								ab
a. b.	the	book	that	shocked	Mary	's	title	sensical nonsensical

6. Left-to-right Quantifier Scope Preference

Let us consider now another instance of ambiguity: quantifier scope preference. A rudimentary account of sentence-peripheral quantifier phrase scoping is obtained in Lambek calculus by means of lexical assignments such as the following:[7]

(38)

$$\begin{array}{lll}
\textbf{everyone} & - & \lambda x \forall y (x\ y) \\
 & := & \mathrm{S}t/(\mathrm{N}\backslash \mathrm{S}t) \\
\textbf{everyone} & - & \lambda x \forall y (x\ y) \\
 & := & (\mathrm{S}t/\mathrm{N})\backslash \mathrm{S}t \\
\textbf{someone} & - & \lambda x \exists y (x\ y) \\
 & := & \mathrm{S}t/(\mathrm{N}\backslash \mathrm{S}t) \\
\textbf{someone} & - & \lambda x \exists y (x\ y) \\
 & := & (\mathrm{S}t/\mathrm{N})\backslash \mathrm{S}t
\end{array}$$

Given these assignments, one analysis of (7a) is that given in Figure 7. This is the subject wide scope analysis: its extracted and simplified semantics are as in (39).

7 For a more refined treatment (not requiring multiple lexical categorizations), for which the results of this paper stand unchanged, see Morrill (1994, Chap. 4).

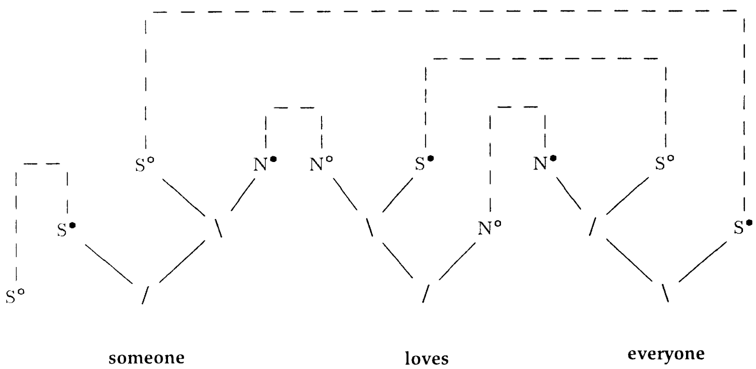


Figure 7
Proof net analysis for the subject wide scope reading of (7a) *someone loves everyone* ($\exists\forall$).

(39) a. $(\lambda x \exists y(x\ y)\ \lambda 1(\lambda x \forall y(x\ y)\ \lambda 2(love\ 2\ 1)))$

b. $\exists x \forall y(love\ y\ x)$

A second analysis is that given in Figure 8. This is the object wide scope analysis: its extracted and simplified semantics are as in (40).

(40) a. $(\lambda x \forall y(x\ y)\ \lambda 2(\lambda x \exists y(x\ y)\ \lambda 1(love\ 2\ 1)))$

b. $\forall y \exists x(love\ y\ x)$

Let us compare the complexity profiles for the two readings:

(41)

4		b		
3			ab	
2		a		
1	ab			
0				ab
a. b.	someone	loves	everyone	$\exists\forall$ (subject wide scope, Figure 7) $\forall\exists$ (object wide scope, Figure 8)

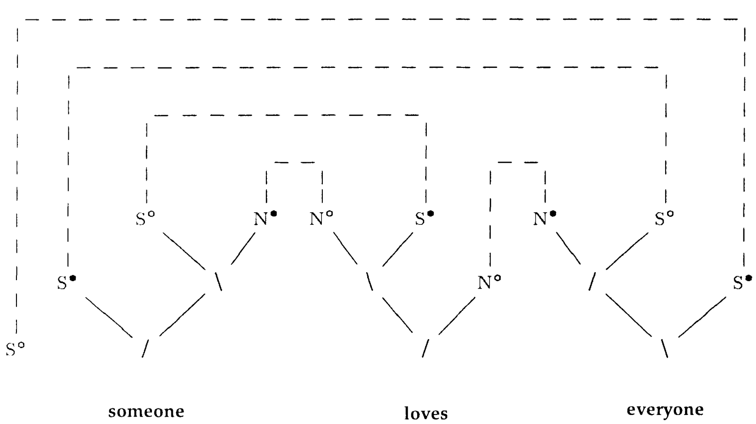


Figure 8
Proof net analysis for the object wide scope reading of (7a) *someone loves everyone* ($\forall\exists$).

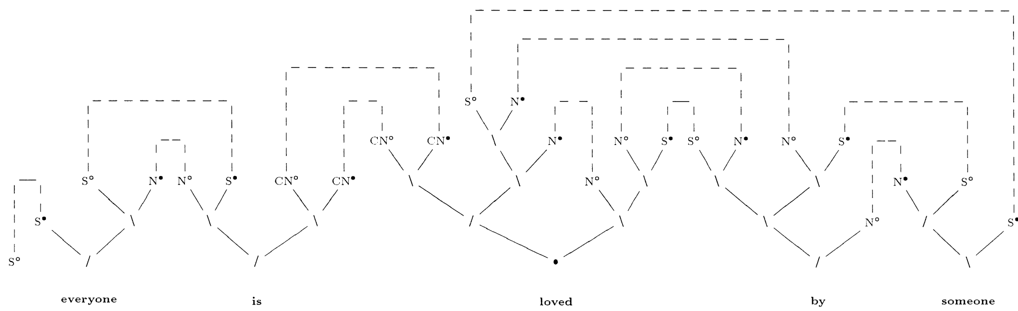


Figure 9
Proof net analysis for the subject wide scope reading of (7b) *everyone is loved by someone* (∀∃).

At the only point of difference, the subject wide scope reading, the preferred reading, has the lower complexity.

We emphasize that some target category is expected at the start of language processing. Lambek proof nets have the property that they can be kept planar under any cyclic permutation of the roots; one can view them as a circular list and require planarity on the interior (or exterior) of a disc. The same proof net would then have different complexity profiles depending on whether one set in axiom links clockwise or counterclockwise. Language processing, though, takes place in time, which is *not* cyclic, so we order the roots along a line, starting, according to the principle of expectation, with the target category. As a reviewer has pointed out, if, for example, the target category were placed at the *end* of the time line, then the predictions of the relative acceptabilities in (41) would be reversed!

For the passive (7b), let there be assignments as in (42). The preposition *by* projects an agentive adverbial phrase; *is* is a functor over (post)nominal modifiers (*the man outside*, *John is outside*, etc.) and passive *loved* is treated exactly like passive *raced* in (27).

(42)

$$\begin{array}{lll}
\textbf{by} & - & \lambda x \lambda y \lambda z[[z = x] \wedge (y\ z)] \\
 & := & ((N\backslash S-)\backslash(N\backslash S-))/N \\
\textbf{is} & - & \lambda x \lambda y(x\ \lambda z[z = y]\ y) \\
 & := & (N\backslash S+)/(CN\backslash CN) \\
\textbf{loved} & - & (\lambda x \lambda y \lambda z[(y\ z) \wedge \exists w(x\ z\ w)], \mathit{love}) \\
 & := & ((CN\backslash CN)/(N\backslash(N\backslash S-)))\cdot(N\backslash(N\backslash S-))
\end{array}$$

A ∀∃ analysis of (7b) is given in Figure 9. This has semantics, after some simplification, as in (43), which is equivalent to (40).

(43) $\forall 16 \exists 9 \exists 7[[9 = 7] \wedge (\mathit{love}\ 16\ 9)]$

An ∃∀ analysis of (7b) is given in Figure 10. This has semantics, after some simplification, as in (44), which is equivalent to (39).

(44) $\exists 16 \forall 14 \exists 7[[7 = 16] \wedge (\mathit{love}\ 14\ 7)]$

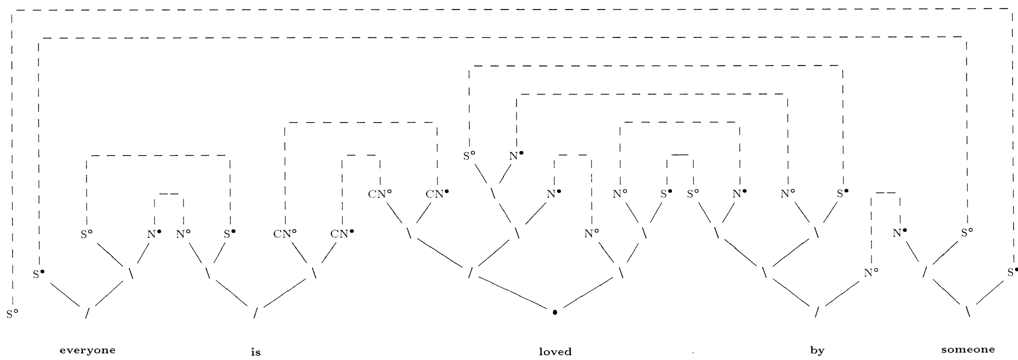


Figure 10
Proof net analysis for the object wide scope reading of (7b) *everyone is loved by someone* (∃∀).

Again, the preferred reading has the lower complexity profile:

(45)

6				b		
5						
4		b	b	a		
3					ab	
2		a	a			
1	ab					
0						ab
a. b.	everyone	is	loved	by	someone	∀∃ (Figure 9) ∃∀ (Figure 10)

7. Preferred Forms

Another good test would be to compare different expressions that are synonymous, holding semantics constant. Our account appears to explain the preference for heavy noun phrases to appear at the end of the verb phrase (heavy noun phrase shift). Of the following two sentences, repeated from (10), the second is more acceptable:

(46) a. ?John gave the painting that Mary hated to Bill.

b. John gave Bill the painting that Mary hated.

The analyses are given in Figure 11. The complexities are thus:

(47)

4							a			
3					a	a		b		
2		ab	ab	a		b	b			
1	ab			b	b			a	a	
0									b	a
a.	John	gave	the	painting	that	Mary	hated	to	Bill	
b.	John	gave	Bill	the	painting	that	Mary	hated		

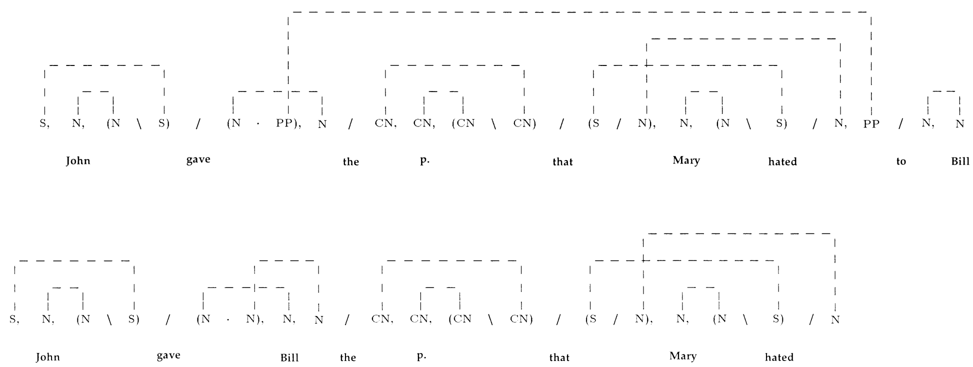


Figure 11
Proof net analysis for (46b) *John gave Bill the painting that Mary hated* (bottom) and (46a) *John gave the painting that Mary hated to Bill* (top).

8. Conclusion: Valencies versus Category Complexity

Finally, another dramatic example of unacceptability is provided by the following:[8]

(48) a. That two plus two equals four surprised Jack.

b. ?That that two plus two equals four surprised Jack astonished Ingrid.

c. ??That that that two plus two equals four surprised Jack astonished Ingrid bothered Frank.

The passive paraphrases, however, seem more or less equally acceptable:

(49) a. Jack was surprised that two plus two equals four.

b. Ingrid was astonished that Jack was surprised that two plus two equals four.

c. Frank was bothered that Ingrid was astonished that Jack was surprised that two plus two equals four.

In Figure 12 we give the analysis of (48b) and in Figure 13 that of (49b). It is very interesting to observe that the complexity profile of the latter is in general lower even though the analysis has more than twice the total number of links.

8 In some linguistic analyses it is claimed that sentencial subjects are obligatorily extraposed into a presentencial topic position, which is not available in nonroot clauses. Under our account no such abstract claims are necessary.

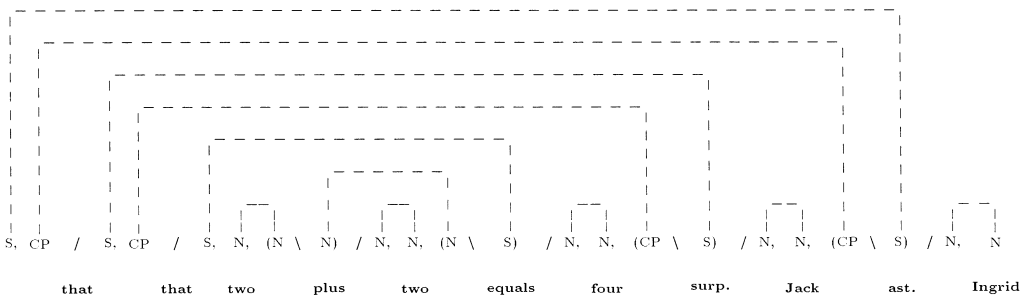


Figure 12
Proof net analysis of (48b) *that that two plus two equals four surprised Jack astonished Ingrid*.

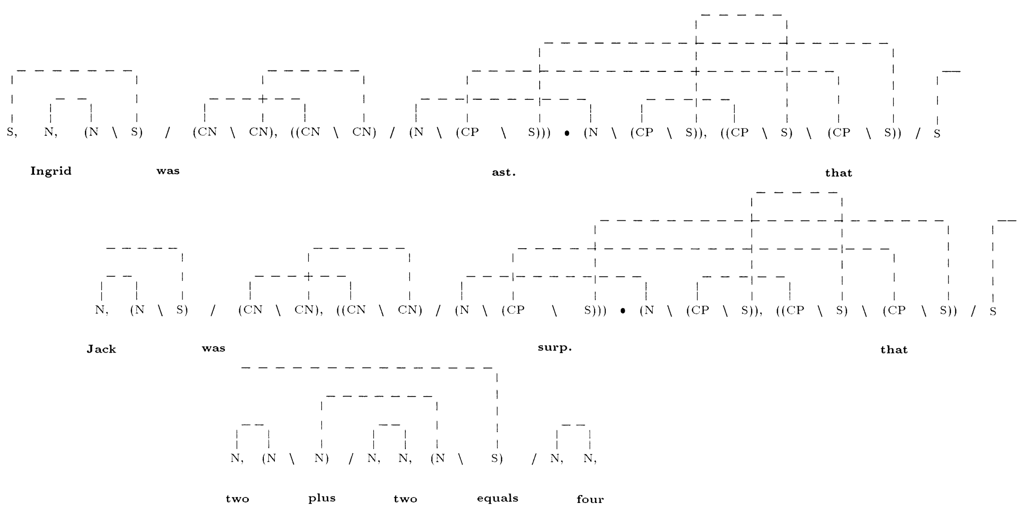


Figure 13
Proof net analysis of (49b) *Ingrid was astonished that Jack was surprised that two plus two equals four*.

(50)

7					a									
6				a		a								
5			a				a							
4				b				ab						
3		a							a		b			
2		b	b			b	b			ab		b		
1	ab				b				b		a		b	
0												a		b
a.	that	that	two	plus	two	equals	four	surp.	Jack	ast.	Ingrid			
b.	Ingrid	was	ast.	that	Jack	was	surp.	that	two	plus	two	equals	four	

Acceptability is no doubt a product of many factors other than just the structural ones considered here. Still, in as much as structural factors may exist, we think this example in particular provides good support for the view that it is resolution of atomic valencies, rather than complexity of categories, which underlies their contribution to acceptability.

Acknowledgments
For helpful remarks I thank Gabriel Bes, anonymous reviewers, and various audiences, and especially for recent comment, Ewan Klein and Mark Steedman.

References

Ades, Anthony E. and Mark J. Steedman. 1982. On the Order of Words. *Linguistics and Philosophy*, 4:517–558.

Bever, Thomas. 1970. The cognitive basis for linguistic structures. In J. R. Hayes, editor, *Cognition and the Growth of Language*, Wiley, New York.

Buszkowski, Wojciech. 1986. Completeness results for Lambek syntactic calculus. *Zeitschrift für mathematische Logik und Grundlage der Mathematik*, 32:13–28.

Carpenter, Bob. 1998. *Type-Logical Semantics*. MIT Press, Cambridge, MA.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA.

Danos, Vincent and Laurent Regnier. 1989. The Structure of Multiplicatives. *Archive for Mathematical Logic*, 28:181–203.

Došen, Kosta. 1992. A brief survey of frames for the Lambek calculus. *Zeitschrift für mathematische Logik und Grundlage der Mathematik*, 38:179–187.

Gibson, E. and J. Thomas. 1996. The processing complexity of English center-embedded and self-embedded structures. *Proceedings of NELS*, University of Massachusetts, Amherst.

Girard, Jean-Yves. 1987. Linear Logic. *Theoretical Computer Science*, 50:1–102.

de Groote, Philippe and Christian Retoré. 1996. On the Semantic Readings of Proof-Nets. In Geert-Jan Kruijff, Glyn Morrill and Richard. T. Oehrle, editors, *Proceedings of Formal Grammar 1996*, Prague, pages 57–70.

Hendriks, Herman. 1993. **Studied Flexibility: Categories and Types in Syntax and Semantics**. D.Phil. dissertation, Universiteit van Amsterdam, Amsterdam.

Hepple, Mark. 1990. Normal form theorem proving for the Lambek calculus. In H. Karlgren, editor, *Proceedings of the 13th International Conference on Computational Linguistics (COLING 1990)*, pages 173–178, Stockholm.

Hepple, Mark and Glyn Morrill. 1989. Parsing and Derivational Equivalence. *Proceedings of the Fourth Conference of the European Chapter of the Association for Computational Linguistics*, pages 10–18, Manchester.

Johnson, Mark. 1998. Proof nets and the complexity of processing center embedded constructions, *Journal of Logic, Language and Information*, 7(4):433–447.

Kimball, John. 1973. Seven principles of surface structure parsing in natural language. *Cognition*, 2:15–47.

König, Esther. 1989. Parsing as natural deduction. *Proceedings of the 27th Annual Meeting*, pages 272–279, Vancouver. Association for Computational Linguistics.

Lamarche, François and Christian Retoré. 1996. Proof nets for the Lambek calculus—an overview. In V. Michele Abrusci and Claudia Casadio, editors, *Proofs and Linguistic Categories: Applications of Logic to the Analysis and Implementation of Natural Language*. CLUEB, Bologna, pages 241–262.

Lambek, Joachim. 1958. The mathematics of sentence structure. *American Mathematical Monthly*, 65:154–170.

Lecomte, Alain and Christian Retoré. 1995. An alternative categorial grammar. Glyn Morrill and Richard. T. Oehrle, editors, *Proceedings of Formal Grammar 1995*, Barcelona, pages 181–196.

Moortgat, Michael. 1988. *Categorial Investigations: Logical and Linguistic Aspects of the Lambek Calculus*. Foris, Dordrecht.

Moortgat, Michael. 1997. Categorial Type Logics. In Johan van Benthem and Alice ter Meulen, editors, *Handbook of Logic and Language*. Elsevier, Amsterdam, pages 93–177.

Morrill, Glyn. 1994. *Type Logical Grammar: Categorial Logic of Signs*. Kluwer Academic Publishers, Dordrecht.

Morrill, Glyn. 1999. Geometry of Lexico-Syntactic Interaction. *Proceedings of the Ninth Conference of the European Chapter of the Association for Computational Linguistics*, Bergen, pages 61–70.

Pentus, Mati. 1994. Language completeness of the Lambek calculus. *Proceedings of the Eighth Annual IEEE Symposium on Logic in Computer Science*, Paris, pages 487–496.

Roorda, Dirk. 1991. **Resource Logics: Proof-Theoretical Investigations**. D.Phil. dissertation, Universiteit van Amsterdam, Amsterdam.

Steedman, Mark J. 1997. *Surface Structure and Interpretation*. MIT Press, Cambridge, MA.

Zielonka, Wojciech. 1981. Axiomatizability of Ajdukiewicz-Lambek calculus by means of cancellation scheme. *Zeitschrift für mathematische Logik und Grundlage der Mathematik*, 27:215–224.