Extraction Covering Extensions of Lambek Calculus are not CF

Martin Emms, Centrum für Informations- und Sprachverarbeitung, Universität München, Germany

Lambek [Lambek, 1958] brought to completion a line of thinking that had been initiated by Ajdukiewicz [Ajdukiewicz, 1935], and carried on by Bar-Hillel [Bar-Hillel, 1953]: that words should be assigned fraction-like categories (that is, built with '/' and '\') and then the categories of phrases deduced from these by universal laws referring to the structure of those categories. Proposals have been made for extending Lambek's framework, increasing the number of possibilities for building categories, and three of these are the concern of this paper: the addition of the so-called *extraction* connective, the addition of the *permutation modality*, and the addition of *universal quantification of category variables*.

In section 1, we consider how adequately Lambek's framework, and the above mentioned extensions, handle *extraction phenomena*. It is a folklore observation about Lambek grammars that they do not permit adequate coverage of extraction. We shall see that there is a narrow and a wide understanding of what it is to give an 'adequate coverage of extraction', and it is only on the wide understanding that the folkloric observation is true. We shall then see that the three extensions of the Lambek calculus enjoy an advantage over the core framework, in allowing coverage of extraction, construed in the wide sense. This is no surprise for the case of the extraction connective and the permutation modality, the arguments for which have usually involved the treatment of extraction. This is more of a surprise in the polymorphic system, the arguments for which have formerly concerned coordination and quantification.

Section 1 shows in a somewhat informal way, that the three extensions of the Lambek calculus increase the possibilities for linguistic coverage. In section 2, we place this statement on a secure mathematical footing, by showing that the recognising power of the extensions of the Lambek calculus exceeds that of grammars based on the Lambek calculus. Section 3 considers how these recognising power results can be carried over to a restricted version of the polymorphic system.

1 Extraction

1.1 The Lambek Calculus

Lambek was concerned with the set of categories generated by the two directions of slash, '/', and '\', and defined a certain set of derivable *sequents* over these categories. This definition is Lambek's calculus, and a sequent generated by the calculus, $x_1, \ldots, x_n \Rightarrow y$ has the intended interpretation that a sequence of expressions having individually the antecedent categories, has as a whole the category of the succedent.

The Lambek Calculus, $\textsc{l}^{(/,\backslash)}$

$$x \Rightarrow x$$

$$\frac{U, x, V \Rightarrow w \quad T \Rightarrow y}{U, x/y, T, V \Rightarrow w}/\text{L} \qquad \frac{T, y \Rightarrow x}{T \Rightarrow x/y}/\text{R}$$

$$\frac{T \Rightarrow y \quad U, x, V \Rightarrow w}{U, T, y\backslash x, V \Rightarrow w}\backslash\text{L} \qquad \frac{y, T \Rightarrow x}{T \Rightarrow y\backslash x}\backslash\text{R}$$

Notational points are that x, y are categories, U, V, T sequences of categories (T possibly empty). x in x/y and $y\backslash x$ will be spoken of as the *value*, and correspondingly, y as the *argument*. The above will be referred to as the Lambek calculus, though strictly speaking, [Lambek, 1958] defines $\textsc{l}^{(/,\backslash,\cdot)}$, a conservative extension of $\textsc{l}^{(/,\backslash)}$, with additional rules governing a further categorial connective, the *product*, $\cdot$. There is some justification for this slight abuse of terminology, because in practice one does not write categorial lexica with a *product* in the value part of the category, and any product appearing in an argument can be eliminated in favour of slashes. It also deserves to be said that nothing in what follows will crucially depend on our making the calculus without product our point of departure: we could just as well proceed on the basis of the calculus with product.

There is some justice to the claim that Lambek grammars are the most general formulation of *bidirectional* categorial grammar. For example, if we drop the Slash Right rules, we obtain what is essentially Bar-Hillel's categorial grammar. Furthermore, the calculus has as *theorems* a number of natural categorial rules which have been argued for at one time or another, for example, Type Raising, Composition and the so-called Geach rule. The final reason for seeing the Lambek calculus as the most general formulation of bidirectional categorial grammar, is that there exist soundness and completeness results for it on a natural interpretation of categories x as sets $[\![x]\!] \subseteq A^*$ of strings over an alphabet A, and connectives interpreted as $[\![x/y]\!] = \{a : a \in A^*,\ a{\cdot}b \in [\![x]\!] \text{ for each } b \in [\![y]\!]\}$, and correspondingly for $\backslash$, where $\cdot$ is concatenation of strings. [Buszkowski, 1982] shows that for this interpretation, we get $\textsc{l}^{(/,\backslash)} \vdash x_1, \ldots, x_n \Rightarrow y$ if and only if $[\![x_1]\!]{\cdot}\ldots{\cdot}[\![x_n]\!] \subseteq [\![y]\!]$ using $\cdot$ as elementwise concatenation.

1.2 Failure to cover extraction with a Lambek Grammar

We will henceforth be concerned with just one aspect of the linguistic coverage that one might seek from a Lambek grammar: coverage of extraction. The following observation has often been made in the Lambek grammar literature ([Moortgat, 1988], [Hepple, 1990], [Morrill et al, 1990]), indicating the Lambek calculus to be a mixture of success and failure:

> *Unbounded extraction from left and right positions can be handled by an* $\textsc{l}^{(/,\backslash)}$*-grammar, but not from medial positions.*

This distinction between left, right and medial extraction is illustrated below:

man who [e likes mary]$_{\mathrm{s}\backslash\mathrm{np}}$	Extraction from left peripheral position
woman who [john likes e]$_{\mathrm{s}/\mathrm{np}}$	Extraction from right peripheral position
dog which [john thinks e bit lucy]$_{?}$	Extraction from non-peripheral position

When the above observation has been made, it has been made as a reflection on the mixed success and failure of a certain strategy towards gaining coverage of extraction. In this strategy, one supposes one has already a Lambek grammar for some extraction-free fragment of English, and then one tries to obtain coverage of extraction simply by adding an appropriately categorised relative pronoun to the lexicon. For this approach to extraction to succeed, there must be a finite set, RC, of categories such that when a string is categorised as s by a Lambek grammar, and an NP is somewhere subtracted, then the resulting string may be categorised by some member of RC. If such a set of categories exists then one can add the relativiser to the lexicon with the category (cn\cn)$/x$, for all x in RC.[1]

This project cannot be carried out in the case of Lambek grammars, because the requisite set RC appears not to exist. While np\s will serve for the case of left-peripheral extraction, and s/np for the case of right-peripheral, it seems no category (or finite set) can be found for the medial cases. Note the problem is not that in such cases of medial extraction, no category can be assigned to the string: there will always be some derivable category for the string. The problem is that there is apparently no category (or finite set of categories) which could be used in all such cases (it is to be noted that the product connective is of no help here).

This then is the folklore concerning extraction and Lambek grammars. As noted above, the observation has been made assuming that one attempts to obtain coverage of extraction in a certain *modular* way, beginning with extraction free constructions, and then simply adding a relative pronoun. If we throw away this assumption, we can in fact quickly show the folklore observation to be *false*, as follows.

We argue basically from the existence of CF PSG's that cover extraction, to the corresponding Lambek grammars. As Gazdar [Gazdar, 1981] has shown, if we begin with a CF-grammar for an extraction-free fragment of English, it is possible to enlarge this grammar to one giving coverage of extraction constructions, in the following way (i) for all the categories x that can dominate an NP, one adds the additional 'slash categories', $x[NP]$, (ii) for any original rule, U, x, $V \Rightarrow y$, where x can dominate an NP, add the rule U, $x[NP]$, $V \Rightarrow y[NP]$, (iii) for any original rule U, NP, $V \Rightarrow y$, add the rule U, $V \Rightarrow y[NP]$ and (iv) add the rule CN, Relpro, $S[NP] \Rightarrow CN$. The resulting grammar will cover extraction constructions, and is subject to no restriction concerning the peripherality of extraction sites. From this CF grammar, using certain well known results, one can argue to the existence of an equivalent Lambek grammar. The main result we need is Gaifman's theorem [Gaifman, 1966], which states that for every CF grammar there is an equivalent, *positive*, Bar-Hillel grammar, where *positivity* is an attribute of a categorial lexicon such that it is easy to see that the language generated on the basis of this lexicon is the same whether the Lambek or Bar-Hillel derivable sequents are assumed as the background calculus. The following notion is used in defining this property:

Definition 1 (Polarity) *The polarity of an occurrence a in the category* $w[a]$ *is defined as follows:*

$$\begin{aligned} pol(w[a]) \quad &= + \text{ if } w[a] = a \\ &= pol(x[a]) \text{ if } w = Un(x[a]), \text{ for some unary category constructor,} \\ &= pol(x[a]) \text{ if } w = Bin(x[a], y), \\ &= \text{opposite of } pol(x[a]) \text{ if } w = Bin(y, x[a]). \end{aligned}$$

The definition is formulated with later extensions of the categorial language in mind. By '$w = Un(x)$' (resp. '$w = Bin(x, y)$') we mean that w is generated from y (resp. x,y) by a *unary* (resp. *binary*) category constructor. The binary constructors we

1. This will certainly allow extraction constructions to be recognised. Whether this will generate also further undesired strings will depend on the initial lexicon.

have met with so far operate as follows '/'$(x, y) = x/y$, '\'$(x, y) = y\backslash x$. We will say that a *connective occurs positively* if it is the main connective of a category that occurs positively, and we will say a grammar is *positive* if all connectives occur positively. Gaifman's theorem provides a positive Bar-Hillel grammar for every CF-PSG, and thereby a Lambek grammar.

So the folklore observation concerning the inadequacy of the Lambek calculus, is in a certain sense, clearly wrong. Despite this, I would like to take the folklore observation nonetheless as pointing us in the right direction, in the direction of an increase to the categorial vocabulary. There are several reasons for this. One reason is that when we apply Gaifman's construction to a Gazdar-style grammar for extraction, the Lambek grammar obtained is opaque in the extreme. By contrast, a categorial grammar obtained by the modular approach would be considerably more perspicuous. The second, more important reason, is that the Gazdar-style grammar which we just described achieves coverage of extraction only if we interpret 'coverage of extraction' in a weak sense. On a more demanding interpretation, that Gazdar-style grammar does not, give coverage of extraction.

To explain this it will be necessary to look a little more closely at what it is to 'cover extraction constructions'. Let us take a 'relative clause' to mean *any* sentence missing an NP. Let us further say that the combination of a determiner, a common noun and a relative clause is a complex NP. One might well take it that to cover extraction constructions is to be able to recognise all complex NP's, as above defined. If so, then the Gazdar-style grammar does not, in this sense, cover extraction constructions. For example, all of the following count as complex NP's, but only the first would be recognised as such by the grammar:

[the a which t_a *believes b likes c]*$_{NP}$ (1 gap)
[the b which [the a which t_a *believes* t_b *likes c]*$_{NP}$ *a-ed]*$_{NP}$ (2 gap)
[the c which [the b which [the a which t_a *believes* t_b *likes* t_c*]*$_{NP}$ *a-ed]*$_{NP}$ *b-ed]*$_{NP}$ (3 gap)

The first counts as complex NP because t_a *believes b likes c* is a relative clause. This implies that *the a which* t_a *believes* t_b *likes c a-ed* is a relative clause, and that therefore the second example is a complex NP. I mean this here in a technical sense, and not that this an acceptable complex NP. In a similar way, we obtain that *[the b which [the a which* t_a *believes* t_b *likes* t_c*]*$_{NP}$ *a-ed]*$_{NP}$ *b-ed* is a relative clause and therefore that the third example is a complex NP (again in a technical sense).

There is a clear sense in which in the first (resp. second,third) complex NP above, there is one (resp. two, three) gaps in the scope of the rightmost relativiser. If we now define an 'n-gap extraction' by the number of gaps that occur in the scope of a relativiser, we can say that the Gazdar-style grammar covers only *1-gap extraction*. So from this grammar, we can conclude that there is a Lambek grammar covering just *1-gap extraction* (though with no peripherality restriction). Furthermore, it is easy to see that if we choose an n, we could generalise the Gazdar-style grammar to cover n-gap extraction, and therefore obtain a Lambek grammar which covers n-gap extraction (again with no peripherality restriction). However, we will never in this way obtain a grammar which for all finite n gives coverage of n-gap extraction. Therefore, if we take coverage of extraction in this rather strong sense, then the folklore observation concerning the inadequacy of the Lambek calculus is actually true.

1.3 Extraction covering extensions of Lambek grammars

At the end of section 1.1, it was claimed that Lambek grammars are the most general formulation of bidirectional categorial grammar. Therefore if we find that Lambek grammars lack some desirable property, our reaction should not be to try to change in some way the set of derivable bidirectional sequents, although this is possible, and has at times been proposed. One might do this for example by adding a previously underivable sequent as an axiom, such as 'disharmonic' function composition, type raising or Geach [Steedman, 1988]. A better response, however, to an observed shortcoming of Lambek grammars, is to *extend the categorial vocabulary*, allowing for a greater variety of categories than those just based on '/' and '\', and then to see whether the categorial grammars based on this extended vocabulary have still the observed shortcoming.

As above argued, Lambek grammars are not able to adequately cover extraction phenomena. We will now describe three different ways to extend the categorial vocabulary and associated rules. All three make it possible to write grammars which cover extraction (in the strong sense described at the end of the previous section). We will use C_1 to C_4 to refer to the calculi obtained by making the following additions to the Lambek calculus:

- C_1: **Arrow** [Moortgat, 1988]

$$\frac{U, y, V \Rightarrow x}{U, V \Rightarrow x \uparrow y}\uparrow\mathrm{R}$$

- C_2: **Permutation Modality** [Morrill et al, 1990]

$$\frac{U,\ x\ \Rightarrow w}{U,\ \Delta x\ \Rightarrow w}\Delta\mathrm{L} \qquad \frac{\Delta T\ \Rightarrow w}{\Delta T\ \Rightarrow \Delta w}\Delta\mathrm{R} \qquad \frac{U,\ y,\ x,\ V\ \Rightarrow w}{U,\ x,\ y,\ V\ \Rightarrow w}\mathrm{Perm},\ \ x \text{ or } y = \Delta z$$

- C_3: **Quantifier** [Emms, 1989]

$$\frac{U, x[y/X], V \Rightarrow w}{U, \forall X.x, V \Rightarrow w}\ \forall\mathrm{L} \qquad \frac{T\ \Rightarrow x}{T \Rightarrow \forall X x[X/Z]}\forall\mathrm{R},\ \ Z \notin FV(T),\ X \notin FV(\forall Z.x)$$

- C_4: **Limitted Quantifier**:
  as for C_3 but with the restriction that only *quantifier-free* categories may be chosen as values for the quantified variable in a ($\forall$L) step.

C_1 is simplification of a system that has been considered by a number of authors recently [Moortgat and Morrill, 1991], [Morrill and Solias, 1993], [Morrill, 1993] and which has its roots in [Moortgat, 1988]. Relative to the Lambek calculus, C_1 has the additional binary category constructor: $\uparrow(x, y) = x \uparrow y$, as the ($\uparrow$R) rule indicates, $x \uparrow y$ is intended to categorise an 'x containing a y gap', that is, a functor which *internally embeds* its argument. It is to be noted that C_1 has no ($\uparrow$L) rule, and on its intended interpretation, it is not possible to formulate such a rule in the kind of sequent calculus we are considering. Sequent calculi with both left and right rules for $\uparrow$ have been suggested, making use of *labels* associated with categories in order to represent the string division operation naturally associated with the $\uparrow$ connective. These calculi contain also rules for a related connective '$\downarrow$', for functors which internally embed themselves into their argument. Let $\mathrm{LDS}^{\uparrow,\downarrow}$ be one of these label-using sequent calculi for $\uparrow$ and $\downarrow$. In the lexica that we will be considering, $\downarrow$ will not occur at all, and $\uparrow$ will only occur *negatively*, with the result that in the possible $\mathrm{LDS}^{\uparrow,\downarrow}$ derivations, ($\downarrow$L), ($\downarrow$R) and ($\uparrow$L) are never operative, and there is always a corresponding C_1 derivation.[2] Therefore what we say concerning C_2 will carry over to these various label-using systems.

2. When $\downarrow$ occurs not at all and $\uparrow$ occurs only *negatively*, then in $\mathrm{LDS}^{\uparrow,\downarrow}$, the term for the consequent is always some permutation, π of the terms for the antecedents, and $\mathrm{LDS}^{\uparrow,\downarrow} \vdash x_1 : u_1, \ldots, x_n : u_n \Rightarrow y : \pi(u_1 u_2 \ldots u_n)$ iff $C_1 \vdash \pi(x_1, \ldots, x_n) \Rightarrow y$

Relative to the Lambek calculus, C_2 extends the categorial vocabulary with the *unary* category constructor, Δ. The main property of Δ is expressed by the Perm rule: that to a category Δx one may apply the structural rule of permutation, one of the structural rules which the Lambek calculus forgoes. C_2 gives the set of rules for Δ that occur in [Morrill et al, 1990]. Since then changes have been argued for [Versmissen, 1992], [Venema, 1993]. The alternative rule sets retain the Permutation rule and ΔL, change the ΔR rule, and in Venema's case also add some additional rules. However, Δ will occur in such a way in the lexica that we will be considering, that only the Lambek rules, the permutation rule and ΔL will have a part to play. Therefore, what we say will carry over to the various other versions of a calculus with a permutation modality.

C_3 is a calculus allowing universal quantification of category variables, and has been proposed as a means to obtain coverage of non-constituent coordination and quantifier scope ambiguity. The basic idea of the ($\forall$R) rule is that the quantified variable should not occur free amongst the antecedents. Beyond the succedents that would be thus derivable, the ($\forall$R) rule also allows the bound variable to be renamed. C_4 is the variant of C_3 that effectively quantifies only over quantifier-free categories, a stipulation that, amongst other things, simplifies the semantic construal of these categories. For further information concerning C_3 and C_4 the reader is referred to [Emms, 1992], [Emms, 1993b] (linguistic applications), [Emms, 1993a] (parsing), [Emms and Leiß, 1993] (Cut elimination).

We will be concerned in this section and the next with C_1, C_2 and C_3. C_4 will considered in the final section. The main observation is that the calculi C_1 to C_3 allow one to find a category or categories for a relative pronoun that will give coverage of extraction. For each of the three calculi, the set of relative pronoun categories will be referred to as rel^i. These rel^i are defined in the following lemma, which states in what sense the various calculi enable coverage of extraction.

Lemma 1 (Extraction for C_i)
if $C_i \vdash U, \mathrm{np}, V \ \Rightarrow t$,
then $C_i \vdash y, U, V \ \Rightarrow \mathrm{cn}\backslash\mathrm{cn}$, *for some* $y \in \mathrm{rel}^i$, *where*

$\mathrm{rel}^1 = \{\mathrm{cn}\backslash\mathrm{cn}/(t \uparrow \mathrm{np})\}$,
$\mathrm{rel}^2 = \{\mathrm{cn}\backslash\mathrm{cn}/(t/\Delta\mathrm{np})\}$,
$\mathrm{rel}^3 = \{\ (\mathrm{cn}\backslash\mathrm{cn})/(t/\mathrm{np}),\ (\mathrm{cn}\backslash\mathrm{cn})/(\mathrm{np}\backslash t),\ \forall X.(\mathrm{cn}\backslash\mathrm{cn})/(X\backslash t)/(X/\mathrm{np})\}$

— U and V not both empty

The fact that the above holds for C_1 and C_2 is no surprise: as extensions of the Lambek calculus, they have been motivated precisely by the coverage of extraction that they permit [Moortgat, 1988], [Morrill et al, 1990]. The fact that the lemma holds for C_3 comes as more of a surprise. As an application of the possibility of quantifying categories, it is rather far removed from the applications concerning coordination and quantifier scope ambiguity. In the case of coordination, the quantified variable is a schematisation over the various categories to be conjoined. In the case of quantification, the quantified variable is an implicit schematisation over the various widths of scope a quantifier might have. In the application to relativisation, the quantified variable is an implicit schematisation over various positions in which a gap could occur.

Proof of Lemma 1: by inspection of the following derivation schemes:

- C_1

$$\dfrac{\mathrm{cn}\backslash\mathrm{cn} \Rightarrow \mathrm{cn}\backslash\mathrm{cn} \qquad \dfrac{U,\ \mathrm{np},\ V\ \Rightarrow t}{U,\ V\ \Rightarrow t\uparrow\mathrm{np}}{\uparrow}\mathrm{R}}{\mathrm{cn}\backslash\mathrm{cn}/(t\uparrow\mathrm{np}),\ U,\ V\ \Rightarrow \mathrm{cn}\backslash\mathrm{cn}}$$

- C_2

$$\dfrac{\mathrm{cn}\backslash\mathrm{cn} \Rightarrow \mathrm{cn}\backslash\mathrm{cn} \qquad \dfrac{\dfrac{\dfrac{U,\ \mathrm{np},\ V\ \Rightarrow t}{U,\ \Delta\mathrm{np},\ V\ \Rightarrow t}\Delta\mathrm{L}}{U,\ V,\ \Delta\mathrm{np}\ \Rightarrow t}\mathrm{Perm}}{U,\ V\ \Rightarrow t/\Delta\mathrm{np}}/\mathrm{R}}{(\mathrm{cn}\backslash\mathrm{cn})/(t/\Delta\mathrm{np}),\ U,\ V\ \Rightarrow \mathrm{cn}\backslash\mathrm{cn}}/\mathrm{L}$$

- C_3

$$\dfrac{\mathrm{cn}\backslash\mathrm{cn}\ \Rightarrow \mathrm{cn}\backslash\mathrm{cn} \qquad \dfrac{\mathrm{np},\ V\ \Rightarrow t}{V\ \Rightarrow \mathrm{np}\backslash t}\backslash\mathrm{R}}{(\mathrm{cn}\backslash\mathrm{cn})/(\mathrm{np}\backslash t),\ V\ \Rightarrow \mathrm{cn}\backslash\mathrm{cn}}/\mathrm{L} \qquad \dfrac{\mathrm{cn}\backslash\mathrm{cn}\ \Rightarrow \mathrm{cn}\backslash\mathrm{cn} \qquad \dfrac{U,\ \mathrm{np}\ \Rightarrow t}{U\ \Rightarrow t/\mathrm{np}}/\ \mathrm{R}}{(\mathrm{cn}\backslash\mathrm{cn})/(t/\mathrm{np}),\ U\ \Rightarrow \mathrm{cn}\backslash\mathrm{cn}}/\mathrm{L}$$

$$\dfrac{\dfrac{\dfrac{\mathrm{cn}\backslash\mathrm{cn}\ \Rightarrow \mathrm{cn}\backslash\mathrm{cn} \qquad V\ \Rightarrow a\backslash t}{\mathrm{cn}\backslash\mathrm{cn}/(a\backslash t),\ V\ \Rightarrow \mathrm{cn}\backslash\mathrm{cn}}/\mathrm{L} \qquad \dfrac{U,\ \mathrm{np},\ V\ \Rightarrow t}{U\ \Rightarrow a/\mathrm{np}}/\mathrm{R},\ n+1 \text{ times}}{\mathrm{cn}\backslash\mathrm{cn}/(a\backslash t)/(a/\mathrm{np}),\ U,\ V\ \Rightarrow \mathrm{cn}\backslash\mathrm{cn}}/\mathrm{L}}{\forall X.(\mathrm{cn}\backslash\mathrm{cn})/(X\backslash t)/(X/\mathrm{np}),\ U,\ V\ \Rightarrow \mathrm{cn}\backslash\mathrm{cn}}\forall\mathrm{L} \qquad a = t/v_n/\ldots/v_1$$

Though it is not my purpose here to argue which of C_1, C_2 and C_3 permits the 'best' treatment of extraction, there is one aspect of the above lemma which does deserve mention. In the case of C_1 and C_2, rel^i consists of a singleton, whereas in the case of C_3, rel^i consists of three categories. Thus it appears the C_3 treatment of extraction is less economical than the C_1 and C_2 treatments. In this respect it is worth mentioning that were we to extend the calculus to one allowing *empty* antecedents, then the corresponding version of the lemma could be proved with rel^3 taken to consist of just the polymorphic relativiser category.

Of relevance for the next section is the generalisation from extraction to topicalisation. For any category x, we can find a 'topicalisation' version of that category x, so that if we move x from somewhere in a derivable sequent to the position of first antecedent, and change x to its topicalisation version, the result is a derivable sequent.

Lemma 2 (Topicalisation for C_i) *For C_1 to C_3,*
if $C_i \vdash U,\ x,\ V\ \Rightarrow t$,
then for some $y \in top^i(x)$, $C_i \vdash y,\ U,\ V\ \Rightarrow t$

$\mathrm{top}^1(x)\ = \{t/(t\uparrow x)\}$,
$\mathrm{top}^2(x)\ = \{t/(t/\Delta x)\}$,
$\mathrm{top}^3(x)\ = \{\ t/(t/x),\ t/(x\backslash t),\ \forall X.t/(X\backslash t)/(X/x)\ \}$
— with U and V not both empty

The proof exactly is exactly parallel to that of the Extraction lemma, and corresponding remarks could be made here concerning whether $top^3(x)$ must contain three categories or could contain just one.

There is a *semantic* side to providing coverage of extraction and we conclude this section with some remarks concerning this. What follows in the next section does not depend on these remarks.

The assignment of meanings to strings is achieved in the case of Lambek grammars by the *term-associated* Lambek calculus. This is an embellishment of the calculus in which sequents are derived featuring category:term pairs, with the intention that variables $u_1, \ldots, u_n$, are associated with antecedents of the final sequent, and some term, Φ, constructed from these variables, is associated with the succedent. This Φ represents in an obvious way the meaning of a compound string, whose constituent words have meanings represented by the variables. There are term associated versions of C_1, C_2 and C_3 and with respect to these we can prove a version of the above Extraction lemma, respecting the standard semantics for relativisation. For the definition of the term-associated Lambek calculus see, for example, [Moortgat, 1988]. The terms are those of typed λ-calculus. For C_1, the term-associated version of ($\uparrow$R) is: *from*, $U,\ y : \zeta,\ V \ \Rightarrow x : \Phi$, *derive* $U,\ V \ \Rightarrow x \uparrow y : \lambda\zeta\Phi$. For the Δ rules of C_2, terms associated with corresponding conclusion and premise occurrences of categories are identical. Finally for C_3, the terms are the terms of 2^{nd} order λ-calculus [Girard, 1972], and the term-associated version of ($\forall$L) is: *from* U, $x[y/X] : u_1,\ V \ \Rightarrow w : \Phi$, *derive* $U,\ \forall X.x : u_2,\ V \ \Rightarrow w : \Phi[u_2(y')/u_1]$, where y' is the type associated with the category y. The term-associated version of ($\forall$R) is not needed for what follows.

The semantic version of the extraction lemma is a matter of defining the *rel*^{i} so that the following holds:

> *if* $C_i \ \vdash U,\ \mathrm{np}{:}x,\ V \ \Rightarrow t{:}\Phi$,
> *then* $C_i \ \vdash y{:}f,\ U,\ V \ \Rightarrow \mathrm{cn}\backslash\mathrm{cn}{:}\Psi$, *for some* $y : f \in \mathrm{rel}^i$, *where*
> $\Psi \rhd \lambda P \lambda y[Py \wedge \Phi[y/x]]$

The term Ψ that we will obtain from the derivation given in the proof of the lemma for C_1 and C_2 will be $f(\lambda x\Phi)$. Thus if for *rel*^{1} and *rel*^{2}, we choose f to be $\lambda Q\lambda P\lambda y[Py \wedge Qy]$, we will obtain the desired reduction property. The same goes for the non-polymorphic derivations in the proof for the lemma for C_3. In the case of the polymorphic derivation, the term Ψ is

$$f((\vec{V}{\rightarrow}t))(\lambda x\lambda\zeta^{\vec{V}}\Phi)(\lambda g^{\vec{V}\rightarrow t} g(\vec{v}^{\vec{V}}))$$

We choose f as $\Delta\pi\lambda Q_1^{e\rightarrow\pi}\lambda Q_2^{\pi\rightarrow t}\lambda P^{e\rightarrow t}\lambda y^e[Py{\wedge}Q_2(Q_1(y))]$. The desired reduction then runs as follows:

$$\Psi = \Delta\pi\lambda Q_1^{e\rightarrow\pi}\lambda Q_2^{\pi\rightarrow t}\lambda P^{e\rightarrow t}\lambda y^e[Py{\wedge}Q_2(Q_1(y))](\vec{V}{\rightarrow}t)(\lambda x^e\lambda\zeta^{\vec{V}}\Phi)(\lambda g^{\vec{V}\rightarrow t} g(\vec{v}^{\vec{V}}))$$
$$\rhd \lambda P^{e\rightarrow t}\lambda y^e[Py \wedge \lambda g^{\vec{V}\rightarrow t}[g(\vec{v}^{\vec{V}})](\lambda x^e\lambda\zeta^{\vec{V}}\Phi(y))]$$
$$\rhd \lambda P\lambda y[y \wedge \Phi[y/x]]$$

2 Recognising Power

The previous section argued that the three extensions of the Lambek calculus can achieve coverage that is beyond the Lambek calculus. Since all are conservative extensions of the Lambek calculus, this gives us, albeit it on a somewhat informal level, a proper inclusion in recognising power between the Lambek calculus and the three extensions. In this section we will prove this inclusion relationship.

To begin with, we note the position of Lambek grammars in the Chomsky heirarchy. Pentus has shown the equivalence between languages recognised on the basis of $\mathrm{L}^{(/,\backslash,\cdot)}$and those recognised by a CF grammar [Pentus, 1992]. This entails an equivalence also for $\mathrm{L}^{(/,\backslash)}$, the Lambek calculus without product. The reverse inclusion is the easy part, being a simple corollary of Gaifman's theorem. For the inclusion from $\textsc{l}^{(/,\backslash)}$-grammars to CF PSG's we argue as follows. An $\textsc{l}^{(/,\backslash)}$ lexicon is already an $\textsc{l}^{(/,\backslash,\cdot)}$ lexicon, but simply one in which the product never appears. On such a lexicon, the language generated by $\textsc{l}^{(/,\backslash)}$ and that generated by $\textsc{l}^{(/,\backslash,\cdot)}$ will not differ, because $\textsc{l}^{(/,\backslash,\cdot)}$ is a conservative extension of $\textsc{l}^{(/,\backslash)}$. Thus we obtain an equivalent CF grammar from an original $\textsc{l}^{(/,\backslash)}$ grammar by simply declaring the original grammar to be a $\textsc{l}^{(/,\backslash,\cdot)}$ grammar, and then proceeding with Pentus' construction. [3]

What we will do in this section is show that C_1, C_2 and C_3 have greater than CF recognising power, and in each case we will exploit the Topicalisation lemma. When one consider topicalisation in natural language it is likely that one only wants to allow certain expressions to topicalise, and it is even more likely that one does not want to allow the process of topicalisation to be iterated. However, in the limitting case of the Topicalisation lemma, we would allow every expression that had a given category x to also have a topicalisation version of x. What we will show now is that when we do this to a language, we obtain at least the permutation closure of the first language. We will later be concerned with particular C_i languages, for which addition of the topicalisation categories will generate exactly the permutation closure.

Theorem 1 *For $i = 1$ to 3, if G_i is a C_i-grammar, there is a C_i-grammar,* top*(G_i) such that* permutation*($L(G_i)$) $\subseteq$ $L($*top*$(G_i))$*

Proof of Theorem 1

We set $top(G_i)$ to be G_i plus *topic* versions of G_i categories

$$G_i = G \cup \{y \to str : x \to str \in G, y \in top^i(x)\}$$

Note that any member of $perm(L(G_i))$, of length n, can be generated from a member of $L(G_i)$ by n *topicalisation* steps, words being fronted in their right to left order in the eventual desired string (and each string moved only once). For example, suppose $abc \in L(G_i)$. Then the various permutations of abc are obtained as follows:

$abc \rightsquigarrow \mathbf{c}ab \rightsquigarrow \mathbf{bc}a \rightsquigarrow \mathbf{abc}$
$abc \rightsquigarrow \mathbf{c}ab \rightsquigarrow \mathbf{ac}b \rightsquigarrow \mathbf{bac}$
$abc \rightsquigarrow \mathbf{b}ac \rightsquigarrow \mathbf{cb}a \rightsquigarrow \mathbf{acb}$
$abc \rightsquigarrow \mathbf{b}ac \rightsquigarrow \mathbf{ab}c \rightsquigarrow \mathbf{cab}$
$abc \rightsquigarrow \mathbf{a}bc \rightsquigarrow \mathbf{ca}b \rightsquigarrow \mathbf{bca}$
$abc \rightsquigarrow \mathbf{a}bc \rightsquigarrow \mathbf{ba}c \rightsquigarrow \mathbf{cba}$

Let us call each such topicalisation step an '*m and m*' topicalisation step, standing for 'move and mark'. We will now show:

> for any string in $L(G_i)$, of length n, there is a categorising sequent for the string that will result after any sequence of 'm and m' topicalisation steps, and this categorising sequent will be of the form $T, U \Rightarrow t$, T being a sequence of p topicalisation categories and U a sequence of $n - p$ G_i categories, p being the number of topicalisation steps.[4]

For each n, the instance of the above for $p = n$, gives that $L(top(G_i)$ contains all permutations of n-length strings in $L(G_i)$. For strings of length 1, the property

3. Note that this result for product free Lambek calculus is not obtained by trying to adapt Pentus' proof: this is blocked because a crucial part of the proof is an interpolation lemma for the calculus with product. It is not known whether the corresponding lemma can be shown for the product-free calculus.
4. One G_i category might already by the topicalisation category of another G_i category. This is not important for our purposes. What we rely on is that every expression with a G_i category has a topicalisation category.

holds trivially, because the string is not changed by an 'm and m' step. So now consider an arbitrexpreary string of length n, $n > 1$. We will show the property for this string by induction on p. Clearly, given the Topicalisation lemma, after 1 'm and m' step, we have a categorising sequent and the categorising sequent is of the form $T, U \Rightarrow t$, with T consisting of 1 *topic* category, and U consisting of $n-1$ G_i categories. Hence we have the property for $p = 1$.

Then suppose up to and including some $p < n$, we have the property, and consider the application of $p+1$ 'm and m' steps to the string. Whatever the $p+1$ 'm and m' steps are chosen to be, let α be the string obtained after the initial p steps. By induction, we have a categorising sequent for α

$T, U \Rightarrow t,$

with T a sequence of p 'topic' categories, and U a sequence of G_i categories, of length $n-p$. At the $(p+1)^{th}$ 'm and m' step, we move and mark some unmarked member of α. Since the first p members of α are marked, we must move and mark the l^{th} member of α, where $l > p$. Because $l > p$, the categorising sequent for α which we assumed by induction can be written as

$T, U_1, x, U_2 \Rightarrow t$

with x the l^{th} antecedent category, and therefore a G_i category. Because x is a G_i category, by the choice of $top(G_i)$, the l^{th} member of α has all members of $top^i(x)$. So by the Topicalisation lemma, for some $y \in top^i(x)$ the string resulting after one more 'm and m' step will have the categorising sequent:

$y,\ T,\ U_1,\ U_2\ \Rightarrow t,$

y, T is a sequence of $p+1$ topic categories, and U_1, U_2 is a sequence of G_i categories, of length $n-p-1$. Hence we have the property for $p+1$

End of Proof

We have shown that adding a topicalisation version of every category gives at least all permutations of all previously recognised strings. It is natural to wonder whether this is exactly the effect. In this regard we note that $L(top(G_i))$ can exceed $permutation(L(G_i))$, an example being:

$$G_i = \left\{ \begin{array}{ll} a & e \\ b & top^i(e)\backslash t \end{array} \right.$$

For each of the C_i we have $C_i \not\vdash e \Rightarrow top^i(e)$, and this gives that $L(G_i)$ is empty. Clearly, however, $L(top(G_i))$ is not empty: it contains ab

We will put the above lemma to use in showing that C_1, C_2 and C_3 all allow the recognition of a certain non-CF language. That language is $permutation((abc)^n)$, and this is not CF, because (i) $a^n b^n c^n$ is not CF, (ii) $a^n b^n c^n = a^* b^* c^* \cap permutation((abc)^n)$, and (iii) CF languages are closed under intersection with regular languages.

In the proof we will make some use of a lemma concerning a certain kind of measure of a sequent called *count* [Benthem, 1988].

Definition 2 (Count) *Where x is any basic category, y any category*

x-count(y) = 1 if $y = x$, = 0 if y is basic and not equal to x
x-count($Un(y)$) = x-count(y)
x-count($Bin(y_1, y_2)$) = x-count(y_1) minus x-count(y_2)
x-count($T \Rightarrow y$) = (sum of x-counts of T) minus x-count(y)

The following can be easily shown by inspection of the proof rules ('$\text{L}^{(/,\backslash)}$+ Perm' is the addition of the C_2 Perm rule to $\text{L}^{(/,\backslash)}$, with all reference to Δ suppressed):

Lemma 3 (Count Invariance for C_1, C_2 and $\text{L}^{(/,\backslash)}$+ Perm) *For $\text{L}^{(/,\backslash)}$ with Perm,* *for C_1 and C_2, we have that when $p_1, \ldots, p_n$ are sequents from which q may be inferred, then the sum of the x-counts of the p_i = x-count(q)*

Given that all derivations end in axioms, and these have zero count, this has the corollary that all derivable sequents of $\text{L}^{(/,\backslash)}$ with Perm, C_1 and C_2 have zero count. The count-invariance property does not hold for C_3 or C_4, because for example s/np, np $\Rightarrow$ s entails $\forall X.\text{s}/X$, np $\Rightarrow$ s, and these two sequents have different np-counts.

Theorem 2 *For i = 1 to 3, there are C_i grammars for the non-CF language, permutation($(abc)^n$)*

Proof of Theorem 2[5]

Let G be the following lexicon:

$$G = \{\ t \rightarrow \Lambda,\ e \rightarrow a,\ (e\backslash s) \rightarrow b,\ (s\backslash t) \rightarrow c,\ (s\backslash(t\backslash t)) \rightarrow c\ \}$$

We first show that $L(G) = (abc)^n$. First note that ab has the category s, and that clearly this is the only string with category s. Second note that both of c's categories have a t value, both have an initial s argument, and no other expression has a t value. Therefore, where x is any string categorised as t, we have

x is categorised as t iff $x = (ab) + c$ or for some y categorised as t, $x = y + (ab) + c$

Therefore $L(G) = (abc)^n$. We now expand G by the topic version of all categories:

$$G_i = G \cup \{y \rightarrow str : x \rightarrow str \in G, y \in top^i(x)\}$$

for example,

$$G_1 \left\{ \begin{array}{llll} \Lambda & t & & \\ a & e & t/(t \uparrow e) & \\ b & (e\backslash s) & t/(t \uparrow (e\backslash s)) & \\ c & (s\backslash t), & t/(t \uparrow (s\backslash t)), & \\ & (s\backslash(t\backslash t)) & t/(t \uparrow (s\backslash(t\backslash t))) & \end{array} \right.$$

$$G_3 \left\{ \begin{array}{lllll} \Lambda & t & & & \\ a & e & t/(t/e) & t/(e\backslash t) & \forall X.t/(X\backslash t)/(X/e) \\ b & (e\backslash s) & t/(t/(e\backslash s)) & t/((e\backslash s)\backslash t) & \forall X.t/(X\backslash t)/(X/(e\backslash s)) \\ c & (s\backslash t) & t/(t/(s\backslash t)) & t/((s\backslash t)\backslash t) & \forall X.t/(X\backslash t)/(X/(s\backslash t)) \\ & (s\backslash(t\backslash t)) & t/(t/(s\backslash(t\backslash t))) & t/((s\backslash(t\backslash t))\backslash t) & \forall X.t/(X\backslash t)/(X/(s\backslash(t\backslash t))) \end{array} \right.$$

We wish to show that $L(G_i) = permutation((abc)^n)$. By Theorem 1, clearly $permutation((abc)^n) \subseteq L(G_i)$. We now show $L(G_i) \subseteq permutation((abc)^n)$.

We need to show that in any recognised string, the number of a's, b's and c's are equal. Define *a-quota*(T) as number of a categories in T. Correspondingly for b and c-quotas. Because the sets of a-categories, b-categories and c-categories are disjoint, we can count the number of a's, b's and c's in a recognised string by the a, b and c quotas of the categorising sequent. Therefore we must show that

for any G_i categorising sequent, $T \Rightarrow t$, the a, b and c quotas are equal.

• G_1 and G_2. G_1 and G_2 give a, b and c more than one category, but the different categories are count-equivalent. In particular:

	e-count	s-count
a-category	1	0
b-category	-1	1
c-category	0	-1

5. The proof follows in several respects one in [Benthem, 1988], which shows that there is an '$\text{L}^{(/,\backslash)}$+ Perm' grammar for *permutation*($(abc)^n$). See final section for relationship to $\text{L}^{(/,\backslash)}$+ Perm

Since $T \Rightarrow t$, our assumed categorising sequent, must have zero x-count, for every possible x-count, we obtain:

e-count(T) = a-quota(T) - b-quota(T) = e-count(t) = 0
s-count(T) = b-quota(T) - c-quota(T) = s-count(t) = 0
$\therefore .a, b$ and c quotas of T are equal.
$\therefore .L(G_i) \subseteq \mathit{permutation}((abc)^n)$, for G_1 and G_2.

• G_3. We cannot reason as we did for G_1 and G_2, as there is no null-count property for C_3 sequents containing quantifiers. Instead we give an argument that:

> if $C_3 \vdash T \Rightarrow t$ then $\mathrm{L}^{(/,\backslash)}$+ Perm $\vdash T' \Rightarrow t$, where T' comes from T simply by replacing all occurrences of $\mathit{top}^3(x)$ by x.

If we have the above, then we can get our necessary identity of quotas for T, from the fact that T' will have the same quotas as T, and the fact that there is a zero-count property for $\mathrm{L}^{(/,\backslash)}$+ Perm.

We begin by noting that as a derivable sequent of C_3, our categorising sequent is also a derivable sequent of C_3 + Perm + Cut. Then if amongst the antecedents, T, there is topic category, $\mathit{top}^3(x)$, we can show that this implies that the corresponding sequent with simply x as an antecedent is also a derivable sequent of C_3 + Perm + Cut.

Observe that for $y \in \mathit{top}^3(x)$, we have C_3 + Perm + Cut $\vdash x \Rightarrow y$, as shown in the following derivations for two of the possible values of $\mathit{top}^3(x)$:

$$\dfrac{\dfrac{\dfrac{\vdots}{t/x,\ x\ \Rightarrow t}}{x,\ t/x\ \Rightarrow t}\text{Perm}}{x\ \Rightarrow t/(t/x)}/\text{R} \qquad \dfrac{\dfrac{\dfrac{\dfrac{\vdots}{X/x,\ x,\ X\backslash t\ \Rightarrow t}}{x,\ X/x,\ X\backslash t,\ \Rightarrow t}\text{Perm}}{x\ \Rightarrow t/(X\backslash t)/(X/x)}/\text{R}/\text{R}}{x\ \Rightarrow \forall X.t/(X\backslash t)/(X/x)}\forall\text{R}$$

Therefore, if the antecedent, T, of our categorising sequent is of the form U, $\mathit{top}^3(x)$, V, then we can infer $U,\ x,\ V\ \Rightarrow t$ is derivable in C_3 + Perm + Cut, simply by a Cut inference:

$$\dfrac{\dfrac{\vdots}{x\ \Rightarrow \mathit{top}^3(x)} \qquad \dfrac{\vdots}{U,\ \mathit{top}^3(x),\ V\ \Rightarrow t}}{U,\ x,\ V\ \Rightarrow t}\text{Cut}$$

We iterate the above reasoning until we obtain some T' which differs from the original T simply by replacing all $\mathit{top}^3(x)$ categories with x. This sequent is derivable in C_3 + Perm + Cut. The argument is then completed as follows:

C_3 + Perm + Cut $\vdash T' \Rightarrow t$
$\therefore .C_3$ + Perm $\vdash T' \Rightarrow t$ (by Cut Elimination for C_3 + Perm [6])
$\therefore .\mathrm{L}^{(/,\backslash)}$+ Perm $\vdash T' \Rightarrow t$ (because T' contains no $\forall$)

End of Proof

The grammar for $(abc)^n$ that we used above was an example of a *positive* grammar. It is easy to show that where all the connectives in T occur positively[Buszkowski, 1988]

> $\mathrm{L}^{(/,\backslash)}$+ Perm $\vdash T \Rightarrow t$ iff for some permutation $\pi(T)$ of T, $\mathrm{L}^{(/,\backslash)} \vdash \pi(T) \Rightarrow t$

---

6. Cut elimination for C_3 + Perm is a small adaption for the Cut Elimination proof for C_3 [Emms and Leiß, 1993]

Given this it is easy to see how the proof of the previous theorem for the case C_3 can be adapted to give a proof of

Theorem 3 *There is a C_3-grammar for the permutation closure of any language recognised by a positive* $\mathrm{L}^{(/,\backslash)}$*-grammar*

As a simple corollary of this and Gaifman's theorem we obtain:[7]

Theorem 4 *There is a C_3 grammar for the permutation closure of any CFG*

The question arises whether one can show a corresponding version of Theorem 3 for C_1 and C_2 grammars, and thereby a corresponding version of Theorem 4. To repeat the proof used for C_3 the essential step will be:

> $C_i \vdash T \Rightarrow t$ implies
> $\mathrm{L}^{(/,\backslash)}$+ Perm $\vdash T' \Rightarrow t$, where T' differs from T simply by the replacement of occurrences of topic categories, by non-topic categories.

We were able to show this implication for C_3 rather easily, with an argument involving Cut and the derivability of topic categories from non-topic categories in C_3 with Perm. We cannot repeat this argument for C_1 and C_2, because when C_i is C_1 or C_2, C_i + Perm $\not\vdash x \Rightarrow y$ for $y \in top^i(x)$. A proof is omitted here, but by a more careful consideration of proof shapes, we can carry out the essential step, and thereby obtain the corresponding version of the above theorems for C_1 and C_2.

3 Limitted Polymorphism

At the outset we introduced C_4, which is that restriction of C_3 that allows only quantifier-free values to be chosen in a ($\forall$L) step. In this section we will see whether this restriction prevents us from proving corresponding versions of the above theorems for C_4.

Recall that the basic building block in the previous section was the Topicalisation lemma, and if we consider this lemma in the case of C_4, we do seem to encounter a significant obstacle: the Topicalisation lemma does not hold without restriction for C_4. The proof of Topicalisation for C_3 relied on the following:

$$\dfrac{\dfrac{\dfrac{\mathrm{cn}\backslash\mathrm{cn} \Rightarrow \mathrm{cn}\backslash\mathrm{cn} \quad V \Rightarrow a\backslash t}{\mathrm{cn}\backslash\mathrm{cn}/(a\backslash t), V \Rightarrow \mathrm{cn}\backslash\mathrm{cn}}/\mathrm{L} \quad \dfrac{U, \mathrm{np}, V \Rightarrow t}{U \Rightarrow a/\mathrm{np}}/\mathrm{R},\ n+1 \text{ times}}{\mathrm{cn}\backslash\mathrm{cn}/(a\backslash t)/(a/\mathrm{np}), U, V \Rightarrow \mathrm{cn}\backslash\mathrm{cn}}/\mathrm{L}}{\forall X.(\mathrm{cn}\backslash\mathrm{cn})/(X\backslash t)/(X/\mathrm{np}), U, V \Rightarrow \mathrm{cn}\backslash\mathrm{cn}}\forall\mathrm{L} \qquad a = t/v_n/\ldots/v_1$$

The value chosen for the quantified variable is

$$t/v_n/\ldots/v_1$$

and should the v_i contain quantifiers, this is a value which could not be chosen in C_4.

7. My thanks here to Makoto Kanazawa, who suggested that Theorem 2 should be generalisable to Theorem 4

We can work our way around this problem in the following way. Consider the quantifiers occurring in V in the provable sequent U, np, $V \Rightarrow t$. The quantifiers in V are constructed in the proof of the sequent, in each each case by a quantifier inference, in the premise of which the quantifier is absent. The thought then arises that under certain conditions, one might be able to build a corresponding proof which simply omits these quantifier steps, and thereby does not build the quantifiers in U, np, $V \Rightarrow t$, building instead U, np, $V' \Rightarrow t$, with V' quantifier free. Then providing we have $V \Rightarrow V'$ ($v_1, v_2 \Rightarrow v_1', v_2'$ is shorthand for $v_1 \Rightarrow v_1'$, $v_2 \Rightarrow v_2'$), we could prove the Topicalisation lemma by choosing as value for the quantified variable of the topicalisation category not t/V, but t/V' ($t/v_1, v_2$ is shorthand for $t/v_2/v_1$).

It turns out that we have this possibility to find a quantifier-free V' such that U, np, $V' \Rightarrow t$ and $V' \Rightarrow V$ are provable when all the quantifiers occur *positively*. To show this we need part (a) of the following (where $\vdash^n$ denotes provability with no more than n steps):

Lemma 4 (Instantiation)
$C_4 \vdash^n \Gamma, x, \Theta \Rightarrow w$ *and* $\forall^-(\Gamma, x, \Theta) = \forall^+(w) = 0$ *implies:*

(a) for some $\forall$*-free* $\overline{x}$, $C_4 \vdash^n \Gamma, \overline{x}, \Theta \Rightarrow w$, $C_4 \vdash x \Rightarrow \overline{x}$

(b) for some $\forall$*-free* $\underline{w}$, $C_4 \vdash^n \Gamma, x, \Theta \Rightarrow \underline{w}$, $C_4 \vdash \underline{w} \Rightarrow w$

Proof of Lemma 4

Without loss of generality, we can restrict attention to proofs with zero-complexity axioms. The proof is then by induction on the sizes of such proofs. For size 1 proofs there is nothing to show. So suppose the lemma (parts a) and b)) for all proofs (ending in zero-complexity axioms) of size less than n, and consider an arbitrary size n proof of $\Gamma, x, \Theta \Rightarrow w$.

a) First consider the case where x is not the active category, then Γ, x, $\Theta \Rightarrow w$, is concluded by some rule R, from some $r_1[x]$, with x as an antecedent, and possibly one further sequent, r_2 (note ($\forall$R) is not amongst the possibilities). See proof (i) below (where $m_1 + m_2 + 1 = n$):

$$\text{i} \quad \dfrac{\dfrac{\vdots m_1}{r_1[x]} \qquad \dfrac{\vdots m_2}{r_2}}{\Gamma,\, x,\, \Theta \Rightarrow w}\,R \qquad\qquad \text{ii} \quad \dfrac{\dfrac{\vdots m_1}{r_1[\overline{x}/x]} \qquad \dfrac{\vdots m_2}{r_2}}{\Gamma,\, \overline{x},\, \Theta \Rightarrow w}\,R$$

Whichever rule R is, it can be confirmed that $\forall^-(r_1[x]) = 0$, and so the inductive hypothesis applies to the left subproof, giving for some $\forall$-free $\overline{x}$ such that $C_4 \vdash x \Rightarrow \overline{x}$, $C_4 \vdash^{m_1} r_1[\overline{x}/x]$. Thus proof (ii) above establishes the claim.

Suppose now that x is the active category. We consider the possibilities for the final step in the proof of Γ, x, $\Theta \Rightarrow w$.

$x = a/b$. We have, where $m_1 + m_2 + 1 = n$, the proof (i) below:

$$\text{i} \quad \dfrac{\dfrac{\vdots m_1}{\Gamma,\, a,\, \Theta \Rightarrow w} \qquad \dfrac{\vdots m_2}{T \Rightarrow b}}{\Gamma,\, a/b,\, T,\, \Theta \Rightarrow w}\,/L \qquad\qquad \text{ii} \quad \dfrac{\dfrac{\vdots m_1}{\Gamma,\, \overline{a},\, \Theta \Rightarrow w} \qquad \dfrac{\vdots m_2}{T \Rightarrow \underline{b}}}{\Gamma,\, \overline{a}/\underline{b},\, T,\, \Theta \Rightarrow w}\,/L$$

The inductive hypothesis applies to the subproofs, so for some quantifier-free $\overline{a}$ and $\underline{b}$, such that $C_4 \vdash$ a $\Rightarrow \overline{a}$, and $C_4 \vdash \underline{b} \Rightarrow$ b, we have the subproofs of (ii). (ii) establishes the claim because $C_4 \vdash a/b \Rightarrow \overline{a}/\underline{b}$

$x = \forall Y.z$. We have the proof (i) below

i $$\dfrac{\dfrac{\vdots n-1}{\Gamma,\ z[y/Y],\ \Theta\ \Rightarrow w}}{\Gamma,\ \forall Y.z,\ \Theta\ \Rightarrow w}\forall L$$

ii $$\dfrac{\vdots n-1}{\Gamma,\ z',\ \Theta\ \Rightarrow w}$$

Because y must be quantifier free, the inductive hypothesis applies to the subproof, so for some quantifier-free z', such that $C_4 \vdash z[y/Y] \Rightarrow z'$, we have the proof (ii) above. (ii) establishes the claim because we have $C_4 \vdash \forall Y.z \Rightarrow z[y/Y]$, and $C_4 \vdash z[y/Y] \Rightarrow z'$, and therefore $C_4 \vdash \forall Y.z \Rightarrow z'$.

b) First consider the case where w is not the active category, then $\Gamma,\ x,\ \Theta\ \Rightarrow w$, is concluded by some rule R, from some $r_1[w]$, with w as the succedent, and possibly one further sequent, r_2. See (i) below (where $m_1 + m_2 + 1 = n$):

i $$\dfrac{\dfrac{\vdots m_1}{r_1[w]} \quad \dfrac{\vdots m_2}{r_2}}{\Gamma,\ x,\ \Theta\ \Rightarrow w}R$$

ii $$\dfrac{\dfrac{\vdots m_1}{r_1[\underline{w}/w]} \quad \dfrac{\vdots m_2}{r_2}}{\Gamma,\ x,\ \Theta\ \Rightarrow \underline{w}}R$$

Whichever rule R is, it can be confirmed that $\forall^-(r_1[w]) = 0$, and so the inductive hypothesis applies, giving for some $\forall$-free $\underline{w}$ such that $C_4 \vdash \underline{w}\ \Rightarrow w$, $C_4 \vdash^{m_1} r_1[\underline{w}/w]$. Thus proof (ii) above establishes the claim.

Now consider the case that w is the active category. There is only the possibility that the last step is (/R), ($\forall$R) being ruled out by the supposition that $\forall^-(\Gamma,\ x,\ \Theta\ \Rightarrow w) = 0$.

$w = a/b$. We have the proof (i) below:

i $$\dfrac{\dfrac{\vdots n-1}{\Gamma, x, \Theta, b \Rightarrow a}}{\Gamma,\ x,\ \Theta\ \Rightarrow a/b}/R$$

ii $$\dfrac{\dfrac{\vdots n-1}{\Gamma, x, \Theta, \overline{b} \Rightarrow \underline{a}}}{\Gamma,\ x,\ \Theta\ \Rightarrow \underline{a}/\overline{b}}/R$$

By applying the inductive hypothesis twice in succession to the subproof of (i) we conclude that for some quantifier-free $\underline{a}$ and $\overline{b}$ such that $C_4 \vdash \underline{a}\ \Rightarrow a$, and $C_4 \vdash b\ \Rightarrow \overline{b}$, we have $C_4 \vdash^{n-1} \Gamma,\ x,\ \Theta,\ \overline{b}\ \Rightarrow \underline{a}$. Hence we have the proof (ii), and this establishes the claim because $C_4 \vdash \underline{a}/\overline{b} \Rightarrow a/b$ $\square$

Given the Instantiation lemma, we can show the following version of the earlier Topicalisation lemma:

Lemma 5 (Topicalisation for C_4)
If $C_4 \vdash U,\ x,\ V\ \Rightarrow t$
then $C_4 \vdash y\ ,\ U,\ V\ \Rightarrow t$ *for some* $y \in top^3(x)$,
— U *and* V *are not both empty*
— $\forall^-(U, x, V \Rightarrow t) = 0$

Proof of Lemma 5

Cases where either U or V are empty are obvious. So suppose

$C_4 \vdash U,\ x,\ V\ \Rightarrow t$, where $\forall^-(U, x, V \Rightarrow t) = 0$.

By the *Instantiation* lemma, we have,

for some quantifier-free $\overline{V}$, $C_4 \vdash V\ \Rightarrow \overline{V}$, $C_4 \vdash U,\ x,\ \overline{V}\ \Rightarrow t$

Hence the following derivation establishes that $C_4 \vdash \forall X.t/(X\backslash t)/(X/x), U, V\ \Rightarrow t$:

$$
\dfrac{\dfrac{t \Rightarrow t \qquad \dfrac{\dfrac{t \Rightarrow t \quad V \Rightarrow \overline{V}}{t/\overline{V},\, V \Rightarrow t}\,1}{V \Rightarrow a\backslash t}}{t/(a\backslash t),\, V \Rightarrow t}\,/\mathrm{L} \qquad \dfrac{U,\, x,\, \overline{V} \Rightarrow t}{U \Rightarrow a/x}}{\dfrac{t/(a\backslash t)/(a/x),\, U,\, V \Rightarrow t}{\forall X.t/(X\backslash t)/(X/x),\, U,\, V \Rightarrow t}\,\forall\mathrm{L}}\,/\mathrm{L}
$$

note: $a = t/\overline{v_n}/\ldots/\overline{v_1}$

End of proof

With this Topicalisation lemma in hand, one can return to the proofs of Theorems 1 to 4, and check that the corresponding statements hold true for C_4.

Corresponding to Theorem 1 we have that for any C_4 grammar, G_4, *not featuring negative quantifiers*, there is grammar $top(G_4)$ generating at least the permutation closure of $L(G_4)$.

This version of Theorem 1 allows us to prove Theorem 2 for C_4: that there is a C_4 grammar for *permutation*$((abc)^n)$. The grammar is obtained by adding the topicalisation versions of all categories in a certain Lambek grammar for $(abc)^n$. The resulting grammar has no negative quantifiers, so that the C_4 version of Theorem 1 can be used to show that at least *permutation*$((abc)^n)$ is recognised. To carry out the part of the proof that establishes that exactly *permutation*$((abc)^n)$ is recognised, we need simply Cut Elimination for C_4 + Perm.

Similarly we obtain versions of Theorems 3 and 4 referring to C_4.

4 Concluding Remarks, Open Questions

On a slightly informal level, we have seen that for C_1 to C_4, there are C_i-grammars that allow give coverage of extraction, something which is not true of the Lambek calculus. Backing this up in a formal way, we have shown that C_1 to C_4 properly exceed the CF recognising power of Lambek grammars, allowing grammars for the permutation closure of a CF language. These facts about the recognising power of C_1 to C_4 are hopefully of interest in their own right. However, I would like to argue that they show up a relationship between coverage of extraction on the one hand, and greater than CF recognising power on the other.

We showed that the C_i allow one to write a topicalisation version of a given category, and that this is simply a slight variation on the categorisation of relative pronouns. We showed that the availability of this topicalisation version entailed the possibility to create permutation closed languages. This was an important half of the proof that there is a C_i-grammar for *permutation*$((abc)^n)$. However, the second half of the proof, which was that the C_i grammars recognised no more than *permutation*$((abc)^n)$, was carried out on a calculus by calculus basis, and did not trace back to the topicalisation lemma. We have therefore in this paper only partly shown that any extraction covering extension of the Lambek calculus will have greater than CF recognising power. Our conjecture, however, is that under a suitable precisation of the relativisation and topicalisation lemmas, we would be able show the connection to greater than CF recognising power quite generally. The precisation of the lemmas should express in some way that the addition of relativisation and topicalisation categories should add nothing more than coverage of relativisation and topicalisation.

A calculus perhaps deserving mention alongside C_1 to C_4 is $\mathrm{L}^{(/,\backslash)}$+ Perm. As for the

C_i calculi, there is a ($L^{(/,\backslash)}$+ Perm)-grammar for the permutation closure of every CF language [Benthem, 1988]. However, unlike the C_i-grammars, ($L^{(/,\backslash)}$+ Perm) grammars do not properly include CF languages. This is because all ($L^{(/,\backslash)}$+ Perm)-grammars are permutation closed, whereas only specially designed C_i-grammars are. This is a side effect of our concern to show that it is possible to obtain permutation closure by more delicate means than simply the presence of a sequent calculus rule to that effect: in the C_i calculi the Perm rule is not admissible. In this respect it is also worth mentioning concerning C_2, the calculus with the permutation modality, that we could have obtained the C_2-grammars from the corresponding ($L^{(/,\backslash)}$+ Perm)-grammars, using the fact that $C_2 \vdash \Delta T \Rightarrow t$ iff $L^{(/,\backslash)}$+ Perm $\vdash T \Rightarrow t$, where ΔT represents a sequence of categories with exactly one Δ and that outermost [Venema, 1993].

The results results reported here are very much a first step in the exploration of the recognising power properties of the various extensions the Lambek calculus mentioned. One further property that has a relatively straightforward proof for other kinds of categorial calculi, seems more problematic in the case of the polymorphic calculi: closure under union. For Lambek grammars, it is relatively easy to prevent the expressions of two languages, L_1 and L_2, from combining: one just distinguishes with subscripts the atomic categories of the two languages. To obtain $L_1 \cup L_2$, is than a matter of taking the lexical entries with a t_1 or t_2 value, and adding an entry with a corresponding t value. For polymorphic grammars, it is not at all evident how to first insulate the expressions of L_1 and L_2 from each other. For example, if an expression of L_1 has an expression categorised $\forall X t_1/X$, this will give a t_1 expression not only when followed by any sequence of L_1 expressions but also when followed by any sequence of L_2 expressions.

Another natural question arising is whether there is a difference in recognising power amongst the C_i-grammars, which we would conjecture to be the case. In a similar vein, one can look for relationships with other grammar formalisms with provably greater than CF-recognising power. In regard to indexed languages, $permutation((abc)^n)$ is conjectured not to be an index-language [Marsh, 1985]. As we have seen it is a C_3-language. An inclusion from index-languages to C_3 languages does not seem likely, however. $a^nb^nc^n$ is an index-language. Whether $a^nb^nc^n$ is a C_3-language is not known, but we would conjecture that it is not.

References

[Ajdukiewicz, 1935] Kazimierz Ajdukiewicz. Die syntaktische Konnexität. *Studia Philosophica*, **1**, 1935, pp1–27. Translated as 'Syntactic Connexion' in Storrs Mc.Call (ed.), *Polish Logic*, 1967.

[Bar-Hillel, 1953] Yehoshua Bar-Hillel. A quasi arithmetic notation for syntactic description. *Language*, **29**, 1953, pp47–58.

[Benthem, 1988] Johan van Benthem. The Lambek Calculus. In *Categorial Grammars and Natural Language Structures*, Oehrle, Bach and Wheeler (eds.), Reidel

[Buszkowski, 1988] Wojczech Buszkowski. Generative Power of Categorial Grammars. In *Categorial Grammars and Natural Language Structures*, Oehrle, Bach and Wheeler (eds.), Reidel

[Buszkowski, 1982] Wojczech Buszkowski. Compatibility of A Categorial Grammar with an associated Category System. In *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, **28**, 1982

[Emms, 1989] Martin Emms. Polymorphic Quantifiers. In Torenvliet, M.S.L. (eds) *Seventh Amsterdam Colloquium Proceedings*, pp139–163

[Emms, 1993a] Martin Emms. Parsing with Polymorphism. In Proceeding of the Sixth Annual Conference of the European Association for Computational Linguistics, 1993

[Emms, 1993b] Martin Emms. Some Applications of Categorial Polymorphism. In DYANA Deliverable R1.3.A, 1993

[Emms, 1992] Martin Emms. Logical Ambiguity. PhD Thesis, Centre of Cognitive Science, Edinburgh.

[Emms and Leiß, 1993] Martin Emms and Hans Leiß. Cut Elimination in the Second Order Lambek Calculus. In DYANA Deliverable R1.1.A, 1993

[Gaifman, 1966] Yeshoshua Bar-Hillel, Chaim Gaifman, E. Shamir. On categorial and phrase structure grammars. *Bull. Res. Council Israel, Sec. F*, **9**, 1–16, 1960.

[Gazdar, 1981] Gerald Gazdar. Unbounded Dependencies and coordinate structure. *Linguistic Inquiry*, **12**, pp155–84

[Girard, 1972] Jean Yves Girard. Interpretation Fonctionelle et Elimination des Coupres de L'Arithmetique d'Order Superieur. PhD Thesis, University of Paris.

[Hepple, 1990] Mark Hepple. *The Grammar and Processing of Order and Dependency: a Categorial Approach.* PhD Thesis, Centre for Cognitive Science, Edinburgh.

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

[Marsh, 1985] W. E. Marsh. Some conjectures on indexed languages. Paper presented to the Association for Symbolic Logic, Stanford University, July 15–19.

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

[Moortgat and Morrill, 1991] Michael Moortgat and Glyn Morrill. Heads and Phrases: Type Calculus for Dependency and Constituency. To appear in *Journal of Language, Logic and Information*

[Morrill, 1993] Glyn Morrill. Discontinuity and Pied Piping in Categorial Grammar. In DYANA Deliverable R1.3.A, 1993

[Morrill et al, 1990] Glyn Morrill, Neil Leslie, Mark Hepple, Guy Barry. Categorial Deductions and Structural Operations. In G. Barry and G. Morrill (eds.) *Studies in Categorial Grammar*, Edinburgh Working Papers in Cognitive Science

[Morrill and Solias, 1993] Glyn Morrill and Teresa Solias. Tuples, Discontinuity and Gapping. In Proceeding of the Sixth Annual Conference of the European Association for Computational Linguistics, 1993

[Pentus, 1992] Mati Pentus. *Lambek Grammars are Context Free.* ms. Dept of Mathematical Logic, Moscow University, 1992

[Steedman, 1988] Mark Steedman. Combinators and Grammars. In *Categorial Grammars and Natural Language Structures*, Oehrle, Bach and Wheeler (eds.), Reidel

[Venema, 1993] Yde Venema. Meeting a Modality. In DYANA Deliverable R1.1.A, 1993

[Versmissen, 1992] Koen Versmissen. Adding a Universal Modality to the Lambek Calculus. In *OTS Yearbook 1992*, Research Institute for Language and Speech, Utrecht University, 1992.