Subsumption in natural language and categorial type logic

Abstract

Existing treatments of subsumption/subtyping are examined and found to be flawed: pure featural decomposition generates unnecessary entailments; typed feature logics beg the question of why disjunction and conjunction of types are needed given that their use overlaps with the subtype hierarchy. An extension of the Lambek calculus is developed which introduces a subtype ordering that generates extra-logical axioms. This results in parsimonious and empirically adequate analyses of, e. g., agreement phenomena while retaining the advantages over feature-based analyses of earlier proposals. The other welcome consequences include a perspective on paradigms that treats them as emergent from syntagmatic classes.

> “*The nature of an appropriate feature system for [Lambek Categorial Grammar] is still an open question.*” —— Bayer & Johnson [BJ95, p. 74]

The question of the nature of an appropriate feature system for Lambek Categorial Grammar is a profoundly uninteresting one, as it wrongly presupposes that feature systems are in and of themselves desirable or even necessary for satisfactory syntactic analyses. Much too often, research in categorial grammar has ignored some of the intricacies of real syntactic phenomena, but those researchers that have taken these challenges seriously have been quick to point out (see for example [Moor88, pp. 9f.] or [Dowt94, §1.6]) that introducing into one's type-logical calculus some mechanism for the featural decomposition of categories would make it possible to deal with phenomena traditionally analyzed by feature-based unification grammars, as for example agreement ([Carp89] presents such an analysis). An important step away from this perspective was made with the work of Bayer/Johnson (surfacing as [BJ95] and as [JB95] at an earlier installment of this conference), who argued that their type-logical treatment of agreement achieved better empirical coverage than any analysis in a unification-based grammar framework. This is due to the fact that categorial type-logics do not focus exclusively on the notion of unification, as many feature logics do, but also incorporate a notion of generalization. Bayer/Johnson's advantage is however not due to their particular encoding of morphological information relevant to agreement. In the rest of this paper we illustrate some of the pleasant consequences of abandoning their unsatisfactory featural decomposition analysis.

1 Subsumption in natural language

For reasons of space, we discuss just one example of a naturally occurring subsumption relation (but will do so in some detail): subject/verb agreement in English. For Bayer/Johnson [JB95, §1] ‘agreement phenomena are modelled in terms of the requirement that arguments must be *subsumed* by, or logically imply, the corresponding argument specification of a predicate’. We couldn't agree more, but nevertheless reject their formalization of the subsumption relation.

The importance of having a notion of subsumption around is obvious: subsumption relations exist in natural languages and must be modeled by a syntactic theory. Consider the case of English subject noun phrases (NPs):

1. Certain NPs can act as subjects of verbs like *is* or *does*; call them ISNPs.
   They include *she* or *the dog*, but not *I*.
2. Certain NPs can act as the subject of the verb *was*; call them WASNPs.
   They include *she*, *the dog*, and *I*. In fact, the ISNPs subsume the WASNPs.
3. Certain NPs can act as subjects of verbs like *are*; call them ARENPs.
   They include *we* and *the dogs*, but not *I*.
4. Certain NPs can act as subjects of verbs like *do*; call them DONPs.
   They include *we*, *the dogs*, and *I*. On a larger scale, they subsume the ARENPs.
5. The WASNPs and the DONPs are not completely disjoint, as both include *I*, which moreover can act as the subject of *am*. Thus we call it an AMNP, which is of course subsumed by both the WASNPs and the DONPs.

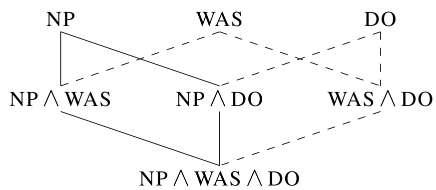
Here we already have an English version of the kind of “overspecification” example that is central to Bayer/Johnson's (and many others') discussion (see [JB95, §6] and references cited there):

(1a)	*He was wrong and do apologize.	(1b)	He/I was wrong.
(2a)	*We was wrong and do apologize.	(2b)	We/I do apologize.
(3)	I was wrong and do apologize.		

2 Previous formalizations of subsumption

The above facts would be analyzed by Bayer/Johnson in terms of some featural decomposition. The exact names of the feature are unimportant; for the sake of simplicity let's say that the WASNPs are analyzed as expressions of the conjunctive type NP $\wedge$ WAS, the DONPs as NP $\wedge$ DO, and *I* consequently as NP $\wedge$ WAS $\wedge$ DO, about which Bayer/ Johnson would remark that such an "inconsistent" type 'may seem ad hoc and unmotivated' [JB95, §6]. However, this only makes sense if the types WAS and DO are incompatible, an assumption not warranted by the facts. The type-logical derivation of (3) is then obvious and proceeds essentially as in [JB95, figure 10].

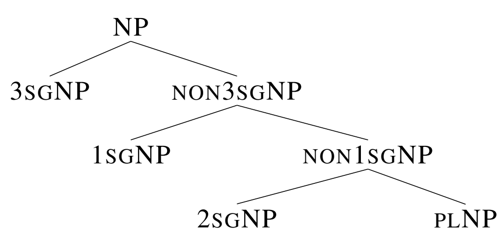
The main problem of a Bayer/Johnson-style analysis is this. The propositional types entailed by NP $\wedge$ WAS $\wedge$ DO form the following meet-semilattice (in fact a Boolean lattice, if one were to add a unit for $\wedge$ on top), but the entailments involving dashed lines are linguistically irrelevant:



(4)

In short, Bayer/Johnson's analysis, though correct, is not very parsimonious, since (the inverse of) the naturally observed subsumption ordering is properly embedded in (i. e., not in one-to-one correspondence with) the logical entailment ordering. This moreover has the methodological disadvantage that in order to get the right entailments one has to first decompose the categories in an arbitrary way: there is no way to get at just the necessary entailments directly without any form of decomposition. Furthermore, a decomposition in terms of conjunctive and/or disjunctive types commits one to the assumption that the subsumption ordering always has a (semi)lattice structure. However we have no reasons to believe that this is in general true of empirically observed sumsumption hierarchies. Other empirical and computational deficiencies of the logic underlying Bayer/Johnson's analysis, and ways to improve its tractability, are discussed in [Moor97, p. 150].

Representative of feature-based grammar frameworks is Sag & Wasow's forthcoming HPSG textbook, which contains an explicit account of the same facts [SW99, §4.6, especially p. 91, fn. 10]. The relevant part of their hierarchy of types is this (after renaming the types and flattening out the AGR-hierarchy into the phrasal hierarchy):



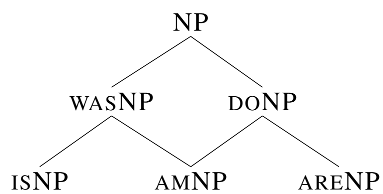
(5)

Unlike the treatment in a Bayer/Johnson-style analysis, Sag & Wasow's subsumption hierarchy is not too large (the 2SGNP and PLNP distinction does exist, but is harder to justify) but too small. It fails to incorporate the insight behind Bayer/Johnson's treatment of "overspecification", which Sag & Wasow would probably call "multiple inheritance": the NP *I*, which belongs to type 1SGNP for Sag & Wasow, is not only a subtype of NON3SG, but also a subtype of those NPs that can be subjects of *was*. What is wrong with Sag & Wasow's story is that the verb *was* has to be specified as selecting the feature-logical disjunction of a 3SGNP and a 1SGNP, whereas a verb like *do* can simply select a NON3SGNP, although the subjects of *do* are not uniform either. Why represent the latter case directly in the type hierarchy but the first as a disjunction of types? It is clear that Sag & Wasow's analysis doesn't live up to the their own expectations: 'The disjunctions needed for describing classes of verbs are thus given by the type hierarchy, not by writing arbitrarily disjunctive lexical entries. One of the goals for a type-based grammar is to predict in this manner which disjunctions will be required for the grammatical analysis of a given language' [SW99, p. 93, fn. 11] (HPSG's "geometrical predictions"). What Sag & Wasow never make fully explicit is how the type hierarchy corresponds to a naturally occurring subsumption relation: one is (isomorphic to) the inverse of the other.

This doesn't mean that a better analysis couldn't be developed within the framework of HPSG. It can, and an important ingredient for this is the proper use of HPSG's type hierarchy, on which very few (but still too many) restrictions are imposed and whose useful aspects can be integrated into a categorial type logic such as the associative Lambek calculus **L** to yield a system we call **L**$<$. We take Sag & Wasow's methodological point to its logical conclusion: the type hierarchy replaces disjunctive and conjunctive types.

3 Subsumption in categorial type logic

We are looking for a way to directly express the observations we made above about the distribution of English subject NPs, illustrated by the following Hasse diagram:



(6)

It so happens that the inverse subsumption relation on English subject NPs forms a join-semilattice and would therefore be best decomposed in terms of disjunctive types, but in general this might not be the case and a decomposition would result in unwarranted entailments. However, all that has to be expressed in terms of entailments is the subsumption relation depicted by this diagram. That is, we want to be able to draw inferences such as AMNP $\vdash$ DONP. In a natural deduction presentation of the Lambek calculus, we would therefore add a rule

$$\frac{a}{b}\ (a < b) \tag{7}$$

where a and b range over atomic types and with the side condition that a be subsumed by b. This requires only that the relation $<$ be an arbitrary relation (specified in the signature) on the set of atomic types.

In a Gentzen calculus, this would correspond to two rules:

$$\frac{a < b \quad \Gamma, b, \Gamma' \vdash C}{\Gamma, a, \Gamma' \vdash C}\ [<\mathrm{L}] \qquad \frac{a < b \quad \Gamma \vdash a}{\Gamma \vdash b}\ [<\mathrm{R}] \tag{8}$$

We retain the assumption that $<$ is defined on atomic types only and also require that it be transitive. Then we prove the following theorem:

1 Theorem (axiomatizability of atomic subtype inference)
The subtyping rules $[<\mathrm{L}]$ and $[<\mathrm{R}]$ and the axiom rule [Ax] can be replaced by a single rule to yield an equivalent deductive system. In the following new rule $\leqslant$ is a reflexive, transitive relation (preorder) on the set of atomic types specified in the signature:

$$\frac{a \leqslant b}{a \vdash b}\ [\mathrm{Ax}\leqslant] \tag{9}$$

2 Proof (sketch only) Subtype inference can be pushed upwards into the leaves. The basic idea is that whenever one finds two steps in a derivation the second of which is subtype inference, the two steps can be interchanged. Two adjacent subtype inference steps are combined into one thanks to transitivity of the subtype relation. An instance of an axiom rule above $[<\mathrm{L}]$ is transformed into another instance of an axiom rule above $[<\mathrm{R}]$:

$$\frac{a < b \quad \dfrac{}{b \vdash b}\,[\mathrm{Ax}]}{a \vdash b}\ [<\mathrm{L}] \quad \rightsquigarrow \quad \frac{a < b \quad \dfrac{}{a \vdash a}\,[\mathrm{Ax}]}{a \vdash b}\ [<\mathrm{R}] \tag{10}$$

It is now easy to construct an algorithm that first pushes $[<\mathrm{L}]$ upwards in a proof tree, merging successive occurrences of $[<\mathrm{L}]$ and transforming them into $[<\mathrm{R}]$ next to an axiom. Then $[<\mathrm{R}]$ is pushed upwards in a similar way. In the end, the proof tree does not contain any subtype inference steps except for possibly single instances of $[<\mathrm{R}]$, which then occur immediately below an axiom rule. $\square$

Using the cut rule, both [$<$L] and [$<$R] can be derived from [Ax$\leqslant$].

Next we extend Lambek's [Lamb58] original cut-elimination proof to the associative Lambek caclulus with atomic subtyping ($\mathbf{L}<$). This easy proof, together with the last result, establishes the decidability of $\mathbf{L}<$.

By adding a preorder on the atomic types we in effect introduce new axioms for the calculus in addition to the identity axiom. Semantically, this is unproblematic, as the usual subset interpretation of $\vdash$ on a standard frame semantics for the Lambek calculus can be kept. If we want to move on to a term calculus, an axiom of the form $a \vdash b$ has to be decorated with an inclusion function $x : a \vdash (\mathsf{a\_to\_b}\ x) : b$ that maps terms x of type a to terms $(\mathsf{a\_to\_b}\ x)$ of type b. In those cases where $a = b$, this would be the identity function on a, which would allows us to write $x : a \vdash x : a$ as usual.

The only previous use of some notion of subtyping in categorial grammar that we are aware of is the use of "prosodic subsorts" in [MO94] and the unpublished work cited there. What [MO94] tried to accomplish is best done in the logical types directly, rather than in the "prosodic labeling".

4 Some additional concepts and consequences

Empty elements and subtyping hierarchies, despite all the heated debate, are formally related in the following trivial way, if we permit sequents with empty antecedents:

$$\dfrac{\dfrac{a < b \quad \dfrac{}{a \vdash a}\,[\mathrm{Ax}]}{a \vdash b}\,[<\mathrm{R}]}{\vdash a\backslash b}\,[\backslash\mathrm{R}] \tag{11}$$

It is advantageous if we don't need to postulate meaningful empty strings. A good example of this is the use of an empty determiner to derive the NP *dogs* from the plural noun *dogs*, which is taken as the basic form. A theory that uses empty determiners would need an additional mechanism to rule out *the and dogs* where *the and* is the conjunction of *the* and the empty determiner. An alternative to this is to have the plural nouns be subsumed by the $\textsc{are}$NPs, either via a lexical rule as in [Carp92] or directly in the syntax through subtyping. Many more of the data in [Carp92] are analyzable in this way. This does not mean that an framework that uses lexical rules, like the one in [Carp92], and one with subtyping are equivalent. Making the plural nouns a subtype of the $\textsc{are}$NPs has certain unwanted side effects; a clearer case where a treatment in terms of multiple lexical entries, which could be related by a lexical redundancy rule, is preferable is this: certain English noun phrases, for example *some fish* and *the sheep*, can act as the subjects of *is* and *are*. However, postulating a common subtype of $\textsc{is}$NP and $\textsc{are}$NP would not be justifiable, since *the sheep* can't be the subject of the conjoined verb phrase *is happy and are grazing*. This example, which seems at first sight similar to *was wrong and do apologize* discussed above, is an instance of a general method for deciding when a subtyping analysis is inappropriate.

Another application of subtyping is the preservation of the integrity of coordinate structures, which has been analyzed in modalized type-logics [Moor97, §4.2]. Assume two clausal types $\textsc{c} \leqslant \textsc{s}$, and say that *and* has types $(\textsc{s}\backslash\textsc{s})/\textsc{s}$ and $((\textsc{c}/\textsc{np})\backslash(\textsc{c}/\textsc{np}))/(\textsc{c}/\textsc{np})$, transitive verbs belong to type $(\textsc{np}\backslash\textsc{c})/\textsc{np}$, main clauses are of type $\textsc{s}$ and relative clauses of type $\textsc{c}/\textsc{np}$ (this is, of course, an oversimplification in many respects: it would have to be extended with a serious account of extraction (see [Morr95] for an overview); the differences and similarities between object and subject NPs can easily be represented by an extended subtype hierarchy). Then *Gottlob admired Kazimierz and Jim detested* [Moor97, p. 114] can have type $\textsc{s}/\textsc{np}$ but not $\textsc{c}/\textsc{np}$, i.e., it could combine with an NP to form a main clause but could not be preceded by a relative pronoun that selects for a $\textsc{c}/\textsc{np}$ type complement.

Our view on subtyping illuminates a further similarity between HPSG and type-logical grammar. In fact, the non-recursive parts of an HPSG type hierarchy can be flattened out and pushed into the phrasal subhierarchy (a procedure mentioned above): for example, if NPs have a $\textsc{head}$-feature $\textsc{case}$ with values *nom* and *acc*, we would create two new NP subtypes *nom_np* and *acc_np*, which would make the feature $\textsc{case}$ dispensable. A more expressive version of subtyping in type-logical grammar, one that lifts the restriction that subtyping only be defined on the atomic types, also answers Sag's challenge (public communication, FHCG '98): how can constructions, analyzed in HPSG in terms of constrained phrasal subsorts, be accomodated in the type-logical framework? Our reply is: by introducing extra-logical axioms into the logics. For example, HPSG's Head-Filler Schema/Rule [SW99, p. 353] would correspond to an axiom like $\textsc{np} \cdot (\textsc{s} \uparrow \textsc{np}) \vdash \textsc{s}$, which looks just like a phrase structure rule (to do this properly, one would have to abstract on $\textsc{np}$ and introduce restricted second-order quantification, or add additional, similar axioms). However, the consequences of introducing arbitrary axioms [Adri92, §3.1.3, pp. 60f.] have to be carefully explored first.

In addition to the methodological (and perhaps pedagogical) advantages of using a type logic with subtyping, we find that it provides a link between syntagmatic information, which could be extracted from a corpus, and paradigmatic information, which is not given *a priori* (at most, it is inherited from traditional grammar, as is partly the case with, e. g., [SW99]) but emerges from the empirically established subtype ordering. Thus among the subject NPs in (6) $\{\mathrm{NP}\}$ is the past tense paradigm; $\{\textsc{is}\mathrm{NP}, \textsc{do}\mathrm{NP}\}$ the present tense paradigm; $\{\textsc{was}\mathrm{NP}, \textsc{are}\mathrm{NP}\}$ the past tense paradigm of *be*; and $\{\textsc{is}\mathrm{NP}, \textsc{am}\mathrm{NP}, \textsc{are}\mathrm{NP}\}$ the present tense paradigm of *be*. More conventionally, the subtype ordering on these NPs corresponds to the subsumption ordering on the verbs that select for them, which can then be classified into their traditional paradigms. What is true of each of the paradigm sets mentioned is that the relation that holds between two types iff they have a common supertype in a given paradigm set is an equivalence relation that partitions the set of minimal elements of the relevant part of the subtype ordering, here the part shown in (6). In fact, this last observation can be distilled into a formal definition of a (disjoint, exhaustive) paradigm.

One question to be answered by further research is how well subtyping can be integrated into more expressive type logics and what the consequences of that would be. What happens if subtyping is extended beyond just atomic types? We conjecture that in this case, cut elimination is still provable, but the two subtyping rule in (8) can definitely no longer be conflated with the axiom rule. What kinds of restrictions will be needed that would keep the resulting logics decidable? Another research avenue has opened up with the presence of the atomic subtyping relation, which is now very similar to a phonological classification system and amenable to quantitative analysis [Broe96]. Finally, the kind of strong correspondence between empirical observations and the theories designed to account for them that we have been insisting on is related to Moshier's goals [Mosh97] of developing a syntactic framework that allows for a radical modularization of a theory, with each modules corresponding to an empirical phenomenon. We answer his question of whether the ideal grammar framework is featureless with 'yes'.

References

[Adri92] Pieter Willem Adriaans. *Language Learning from a Categorial Perspective*. Ph.D. thesis, Universiteit van Amsterdam, 1992.

[BJ95] Sam Bayer and Mark Johnson. Features and agreement. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, pp. 70–76. San Francisco: Morgan Kaufmann, 1995.

[Broe96] Michael Broe. A generalized information-theoretic measure for systems of phonological classification and recognition. In *Computational Phonology in Speech Technology, Proceedings of the Second Meeting of the ACL Special Interest Group in Computational Phonology*, edited by Richard Sproat. Santa Cruz, CA, 1996.

[Carp89] Bob Carpenter. *Phrase Meaning and Categorial Grammar*. Ph.D. thesis, University of Edinburgh, 1989.

[Carp92] Bob Carpenter. Categorial grammars, lexical rules and the English predicative. In *Formal Grammar: Theory and Implementation*, edited by Robert Levine, vol. 2 of *Vancouver Studies in Cognitive Science*, pp. 168–242. New York: Oxford University Press, 1992.

[Dowt94] David Dowty. The role of negative polarity and concord marking in natural language reasoning. In *Proceedings of SALT IV*, edited by Mandy Harvey and Lynn Santelmann. Ithaca, NY: CLC Publications, Cornell University, 1994.

[JB95] Mark Johnson and Sam Bayer. Features and agreement in Lambek categorial grammar. In *Proceedings of the 1995 Formal Grammar Workshop*. Barcelona, 1995.

[Lamb58] Joachim Lambek. The mathematics of sentence structure. *American Mathematical Monthly* 65: 154–170, 1958.

[MO94] Michael Moortgat and Richard Oehrle. Adjacency, dependency and order. In *Proceedings of the Ninth Amsterdam Colloquium*, edited by Paul Dekker and Martin Stokhof. ILLC, Universiteit van Amsterdam, 1994.

[Moor88] Michael Moortgat. *Categorial Investigations: Logical and Linguistic Aspects of the Lambek calculus*. Ph.D. thesis, Universiteit van Amsterdam, 1988.

[Moor97] Michael Moortgat. Categorial type logics. In *Handbook of Logic and Language*, edited by Johan van Benthem and Alice ter Meulen, chap. 2, pp. 93–177. Amsterdam: Elsevier, 1997.

[Morr95] Glyn Morrill. Discontinuity in categorial grammar. *Linguistics and Philosophy* 18: 175–219, 1995.

[Mosh97] M. Andrew Moshier. Is HPSG featurless or unprincipled? *Linguistics and Philosophy* 20: 669–695, 1997.

[SW99] Ivan A. Sag and Thomas Wasow. *Syntactic Theory: A Formal Introduction*. Stanford, CA: CSLI, draft of January 20, 1999, to appear Spring 1999.