Welsh Prenominal Numerals
Ingo Mittendorf and Louisa Sadler[1]

1 Numeral Constructions

Welsh has two numeral-noun-constructions (as well as two counting systems, an older vigesimal system and a modern decimal system). One involves a (pseudo-)partitive construction in which a numeral is followed by *o* 'of' and a plural noun, as shown in (1). In this construction, the numeral can take wide scope over a coordination, and the construction shows no particular syntactic or morphological pecularities.

(1) pump o fechgyn
five of boy.PL
'five boys'

(2) pump o fechgyn a merched
five of boy.PL and girl.PL
'five boys and girls'

The other (Num+N) construction consists of a numeral directly followed by a *singular* noun.

(3) pum bachgen
five boy.SG
'five boys'

(4) pum ci
five dog.M.SG
five dogs

(5) tair cath
three.F cat.F.SG
three cats

This construction shows a number of pecularities.

1.1 Welsh NP Structure

Determiners and a small and very restricted set of Adjectives occur prenominally. Most adjectives are postnominal, as are demonstratives (which co-occur with the definite article). The definite article *y* has the (purely positional) variants *yr* and *'r*. It does not itself distinguish either number or gender. There is no indefinite article. Demonstratives agree in GEND and NUM and follow any postnominal Adjectives. In the modern language, only a few adjectives have distinct FEM, MASC forms, and with these the use of the feminine forms is mostly optional in the modern language. Amongst numerals, only the paucals 2-4 have distinct FEM, MASC forms.

(6)

MASC	FEM		MASC	FEM	
gwyn	*gwen*	white	*cryf*	*cref*	strong
melyn	*melen*	yellow	*trwm*	*trom*	heavy
bychan	*bechan*	small	*byr*	*ber*	short

(7)

MASC	FEM	
dau	*dwy*	2
tri	*tair*	3
pedwar	*pedair*	4

Words in Welsh are subject to initial consonant mutation (ICM) in a variety of lexical and syntactic contexts. FSG NPs are soft-mutated after the definite article *y*. Postnominal APs following a FSG noun are soft-mutated.

(8) torth fawr (< mawr)
loaf.F.SG big
a big loaf

(9) torthau mawr
loaf.F-PL big
big loaves

[1]This work was carried out with the support of the AHRC under grant AH/E006868/1, which we gratefully acknowledge.

1.2 Num+N Construction

Postnominal adjectives (e.g. *arall*) agree with the N as obligatorily singular with a prenominal numeral.

(10) cŵn eraill
dog.M.PL other.PL
other dogs

(11) pum ci arall
five dog.M.SG other.SG
five other dogs

Article and demonstratives are plural.

(12) y gath hon
the cat.F.SG this.F.SG
this cat

(13) y tair cath hyn
the three cat.F.SG this.PL
these three cats

Note ICM contrast: *cath* in (12) is soft-mutated after the definite article because the N is FSG but *tair* in (13) does not show soft mutation (*tair* > *dair*) because the N is treated as PL (although morphologically SG).

Externally, the NP is PL in terms of other agreement processes (e.g. pronominal anaphora).

(14) Roedd y pum dyn yn gweld eu hunain yn y drych.
be.IMPERF.3S the five man.M.SG PT see 3PL self in the mirror
The five men saw/were seeing themselves in the mirror.

(15) Cafodd y pum ci eu curo.
get.IMPERF.3S the five dog.M.SG 3PL beat
The five dogs were beaten

Note that in Welsh it is not the case that the singular is a morphological form unmarked for number which can be widely used wherever plurality can be deduced from the context: it is generally impossible to use a singular to denote an aggregate.

Scope over a coordination of N appears to be impossible in the Num+N construction (for most speakers).

(16) *pum bachgen a merch
five boy.SG and girl.SG
'five boys and girls'

Broadly what this means is that a sentence like the following is in general grammatical only under the reading in which the numeral scopes only over the first nominal.

(17) Dinistriwyd pedwar tŷ a bwthyn gan y storm.
was-destroyed four house and cottage by the storm
∗Four dwellings (houses and cottages) were destroyed by the storm
∗Four houses and four cottages were destroyed by the storm
Four houses and a cottage were destroyed by the storm

Why is it not possible for a prenominal numeral to scope over a coordination of nouns?

2 N N Coordination

There does not seem to be a reason to rule this out in principle and it is certainly possible to coordinate V at the lexical level. An adjective (or a strong form determiner) can scope over a coordination on nouns, hence N coordination must be permitted.

(18) pa unigolion a sefydliadau
which individuals and institutions

(19) y gwahanol afiechydon a chlefydau
the different illnesses and diseases

(20) pa syniadau a chysyniadau
which ideas and concepts

(21) pob athro ac athrawes
every teacher.MSG and teacher.FSG

Note however that prenominal material such as the definite article and the proclitic possessors must be repeated.

(22) a. y dynion a'r merched *vs.* *y dynion a merched
the men and=the girls vs. *the men and girls
'the men and girls' PWT p. 265

b. ei ffagots a'i phys hi
CL.3SG.F faggots and=CL.3SG.F peas PRON.3SG.F
'her faggots and peas' PWT p. 209

The assumption that you can get N coordination might just be wrong: natural coordinations?

3 Previous Treatment of Num/N Combinations

Mittendorf and Sadler (2005) assume the following (taking NUMB as a type of ADJUNCT):

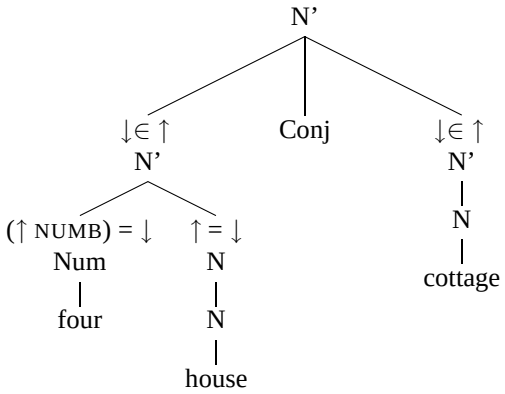
(23) a. *tŷ* N (↑ CONC NUM) = SG
{ (↑ INDEX NUM) = SG
| (↑ INDEX NUM) =$_c$ PL }
b. *tai* N (↑ CONC NUM) = PL
(↑ INDEX NUM) = pl
c. *pedwar* NUM ((NUMB ↑) CONC NUM) =$_c$ SG
((NUMB ↑) INDEX NUM) = PL

(24)

	Det	Num	N	Adj	Demon
IND	PL	**PL**			PL
CONC			**SG**	SG	

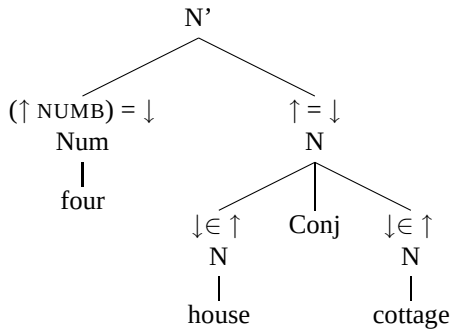
This appears to be over-permissive. It accounts for the grammatical reading of (25).

(25) [pedwar tŷ] a [bwthyn]
four house.SG and cottage.SG
'four houses and a cottage'



But also the following depending on whether the NUMB is a distributive GF/property.

(26) pedwar [tŷ a bwthyn]
four house.SG and cottage.SG
'four houses and cottages'



If NUMB is distributive the constraints above are satisfied in the following structure.

(27)
$$\begin{bmatrix} \text{CONJ} & \text{AND} \\ \text{IND} & \begin{bmatrix} \text{NUM} & \text{PL} \end{bmatrix} \\ \left\{ \begin{bmatrix} \text{PRED} & \text{'HOUSE} \\ \text{IND} & \begin{bmatrix} \text{NUM} & \text{PL} \end{bmatrix} \\ \text{CONC} & \begin{bmatrix} \text{NUM} & \text{SG} \end{bmatrix} \\ \text{NUMB} & \begin{bmatrix} \text{PRED} & \text{'FOUR'} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \text{PRED} & \text{'COTTAGE'} \\ \text{IND} & \begin{bmatrix} \text{NUM} & \text{PL} \end{bmatrix} \\ \text{NUMB} & \begin{bmatrix} \text{PRED} & \text{'FOUR'} \end{bmatrix} \\ \text{CONC} & \begin{bmatrix} \text{NUM} & \text{SG} \end{bmatrix} \end{bmatrix} \right\} \end{bmatrix}$$

If NUMB is non-distributive then the constraints above are again satisfied.

(28)
$$\begin{bmatrix} \text{CONJ} & \text{AND} \\ \text{IND} & \begin{bmatrix} \text{NUM} & \text{PL} \end{bmatrix} \\ \text{NUMB} & \begin{bmatrix} \text{PRED} & \text{'FOUR'} \end{bmatrix} \\ \left\{ \begin{bmatrix} \text{PRED} & \text{'HOUSE} \\ \text{IND} & \begin{bmatrix} \text{NUM} & \text{SG} \end{bmatrix} \\ \text{CONC} & \begin{bmatrix} \text{NUM} & \text{SG} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \text{PRED} & \text{'COTTAGE'} \\ \text{IND} & \begin{bmatrix} \text{NUM} & \text{SG} \end{bmatrix} \\ \text{CONC} & \begin{bmatrix} \text{NUM} & \text{SG} \end{bmatrix} \end{bmatrix} \right\} \end{bmatrix}$$

However it seems that in general speakers do not permit either of these, interpreting the string only as in (25).

3.1 Natural Coordination Structures

One can wonder hypothetically about how the current proposal for Num Noun combinations and the current proposal for Natural Coordination would combine: *y pum nod ac amcan* "the five aims and objectives", which an informant accepted, may be such an example. Dalrymple and Nikolaeva (2006) propose that such natural coordination structures involve "resolution" of both the IND and the CONC NUM value to PL. The constraints on the numeral will be interpreted with respect to the f-structure (f) and will fail.

(29)
$$f\begin{bmatrix}\text{CONJ} & \text{AND}\\ \text{IND} & [\text{NUM} \quad \text{PL}]\\ \text{CONC} & [\text{NUM} \quad \text{PL}]\\ \text{NUMB} & [\text{PRED} \quad \text{`FOUR'}]\\ \text{CONJ1} & \begin{bmatrix}\text{PRED} & \text{`HOUSE}\\ \text{IND} & [\text{NUM} \quad \text{SG}]\\ \text{CONC} & [\text{NUM} \quad \text{SG}]\end{bmatrix}\\ \text{CONJ2} & \begin{bmatrix}\text{PRED} & \text{`COTTAGE'}\\ \text{IND} & [\text{NUM} \quad \text{SG}]\\ \text{CONC} & [\text{NUM} \quad \text{SG}]\end{bmatrix}\end{bmatrix}$$

The following was graded with a high degree of acceptability by nearly all respondents (but may be a lexicalised binominal and hence irrelevant), but it might suggest rethinking the representation of natural coordinations for Welsh.

(30) Mae'na bum cwpan a soser ar hugain yn y cwpwrdd.
is-there five cup and saucer on twenty in the cupboard
There are 25 cups and saucers in the cupboard

4 Possible Analyses

4.1 Brute Force

We constructed questionnaire data designed to disambiguate the examples and explicitly probe for either the quantified (union) or the distributive reading. The likelihood is that for most speakers NUMB is not a distributive GF but that the wide scope numeral is also pretty bad under union readings.

Assuming that NUMB is not distributive, then a syntactic account could be given by altering the constraints to the following.

(31) *tŷ* N (↑ CONC NUM) = sg
{ (↑ INDEX NUM) = sg
¬ (NUMB (∈) ↑)
| (↑ INDEX NUM) =$_c$ pl
(NUMB ↑) }

4.2 Interaction with CCA

It might be that this comes about because there is an interaction with CCA. Welsh predicate-argument agreement involves CCA in INDEX. A possible explanation for why (28) is unacceptable might be the following. The

agreement constraint associated with the numeral is enforced on the closest conjunct and hence the f-structure is in fact the following.

(32)
$$\begin{bmatrix} \text{CONJ} & \text{AND} \\ \text{IND} & \begin{bmatrix} \text{NUM} & \text{PL} \end{bmatrix} \\ \text{NUMB} & \begin{bmatrix} \text{PRED} & \text{`FOUR'} \end{bmatrix} \\ \left\{ \begin{bmatrix} \text{PRED} & \text{`HOUSE} \\ \text{IND} & \begin{bmatrix} \text{NUM} & \text{PL} \end{bmatrix} \\ \text{CONC} & \begin{bmatrix} \text{NUM} & \text{SG} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \text{PRED} & \text{`COTTAGE'} \\ \text{IND} & \begin{bmatrix} \text{NUM} & \text{SG} \end{bmatrix} \\ \text{CONC} & \begin{bmatrix} \text{NUM} & \text{SG} \end{bmatrix} \end{bmatrix} \right\} \end{bmatrix}$$

The idea would be that given this structure, there is some sort of clash which gets in the way of interpretation - however it is difficult to see why such a clash is a problem in this case and not e.g. in the case of subject predicate and other agreement, which is enforced on the one conjunct without leading to any interpretation problems. More radically, suppose that NUMB itself is not distributed but is contributed *only* to the first conjunct. Then what actually results is the following (and this is the f-structure which corresponds to the *four houses and a cottage reading*.....).

(33)
$$\begin{bmatrix} \text{CONJ} & \text{AND} \\ \text{IND} & \begin{bmatrix} \text{NUM} & \text{PL} \end{bmatrix} \\ \left\{ \begin{bmatrix} \text{PRED} & \text{`HOUSE} \\ \text{NUMB} & \begin{bmatrix} \text{PRED} & \text{`FOUR'} \end{bmatrix} \\ \text{IND} & \begin{bmatrix} \text{NUM} & \text{PL} \end{bmatrix} \\ \text{CONC} & \begin{bmatrix} \text{NUM} & \text{SG} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \text{PRED} & \text{`COTTAGE'} \\ \text{IND} & \begin{bmatrix} \text{NUM} & \text{SG} \end{bmatrix} \\ \text{CONC} & \begin{bmatrix} \text{NUM} & \text{SG} \end{bmatrix} \end{bmatrix} \right\} \end{bmatrix}$$

4.3 Non-Projecting Words

Asudeh (2002) proposes an analysis of Celtic preverbal particles as non-projecting words (Toivonen, 2003), but it is not clear that treating prenominal numerals as $\hat{\text{N}}$um will help. Since the numeral can in fact be complex, how would it be represented as the non-projecting part? Moreover a non-projecting element can normally (?) take syntactic scope. The natural way to force the nominal not to be a coordinate structure might be to make the N sister of Num a non-projecting word, but this is not consistent with the principles for labelling the mother node, which we would expect to be N rather than NumP.

4.4 Numeral with Compromised Autonomy

The prenominal numeral does not have the status of an affix but a complete word. They can themselves undergo further affixation to form ordinal numbers (so are more stem-like than affix-like). They form their own domain with respect to word-level morphophonology. Outside of this construction, they can stand alone in argument positions.

The default assumption would appear to be that they are, to all intents and purposes, complete morphological words. It is still possible that they participate in some word level morphological processes such as compound formation. If they do, then we would expect, according to BM95, the resultant structures to be syntactically opaque. Indeed to illustrate that morphological structures are not accessible to syntactic processes such as extraction, conjunction, gapping, inbound anaphora, Bresnan and Mchombo (1995) illustrate the bar on extraction with *American history teachers* which they clearly therefore take to be a compound.

Moreover the behaviour wrt to coordination suggests a lack of syntactic autonomy.

(34) If an item must be repeated on each conjunct in a coordinate structure, then it must be an affix and cannot be a PLC. [= postlexical clitic]
If an item must fail to be repeated on each conjunct in a coordinate structure, then it must be a PLC and cannot be an affix.
"Both of these criteria follow immediately from the definitions of affixes and PLCs and from any theory which allows for constituent coordination [...]" (Miller, 1992, 386)

4.5 Lexical Sharing

Lexical Sharing involves having the λ function give two adjacent nodes a single exponent, that is a lack of morphological autonomy conbined with an adjacency constraint over nodes. *Typical cases must then have some evidence for autonomous nodes in the syntax, otherwise they would be better treated purely morphologically*.[2]

(35)

Turkish suspended affixation	English nonsyllabic Aux contraction
Romance/German P-Det contractions	Udi Person Markers

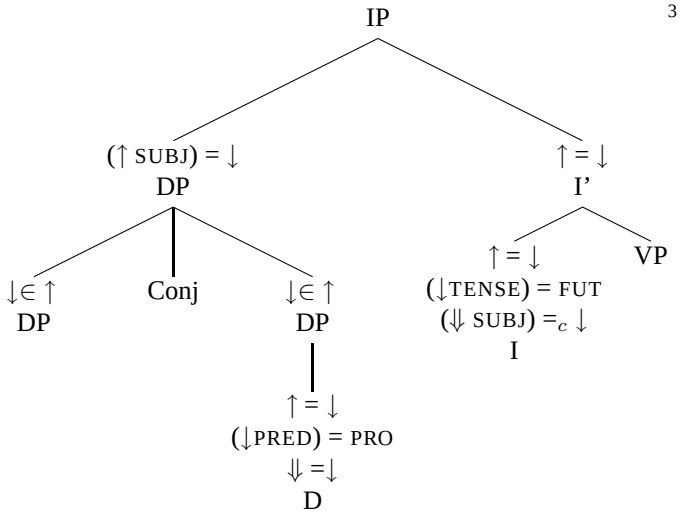
(36)

I'll help	non-syllabic/syllabic
She and I'll help	syllabic only
I'll be there on Sunday and am looking forward to meeting you	non-syllabic/syllabic

$\Downarrow$ abbreviates $\phi(\lambda(*))$ 'the f-structure of the exponent of the current node'

(37)

I'll	$\leftarrow$	D	I
		$\Downarrow = \downarrow$	$(\Downarrow \text{SUBJ}) =_c \downarrow$
		$(\downarrow \text{PRED}) = \text{PRO}$	$(\downarrow \text{TENSE}) = \text{FUT}$



[3]

[2]Westcoat suggests that "the conjunction in [(36c)] suggests that it corresponds to two constituents, a D and an I; the auxiliary I resides in the left-hand conjunct, while the pronominal D lies outside the conjunction and takes scope over both conjuncts. Thus non-syllabic auxiliary constractions are good examples of lexical sharing"(Wescoat, 2000).

[3]Formulation in 2005 paper is slightly different:

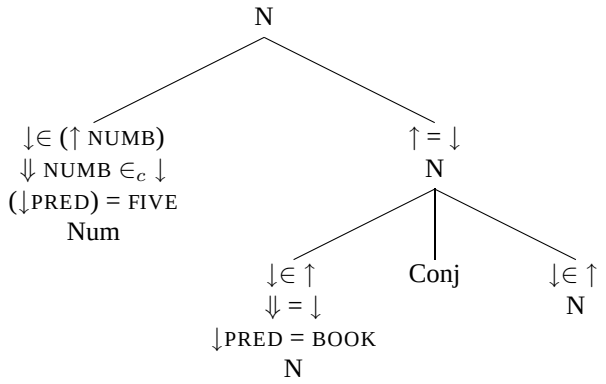
A lexical sharing spell out rule for the Num Noun combination would look like this:

(38)

N	⟶	N		CONJ		N
		↓ ∈ ↑				↓ ∈ ↑
N	⟶	Num			N	
		↓ ∈ (↑ ADJ)			↑ = ↓	

(39)

pum llyfr	←	Num	N
		(⇓ NUMB) ∈ $_c$ ↓	⇓ = ↓



With NUMB non-distributive, the constraining equation (⇓ NUMB) ∈ $_c$ ↓would fail to be satisfied.

- the treatment seems to hinge on the use of a constraining membership statement
- lexical sharing seems appropriate where one string occupies two nodes, here if anything we arguably have the reverse situation

4.6 Transparent Compounds?

On balance, however, it does not seem persuasive to treat the Num N combination as a morphological compound (and hence unanalysed whole in the c-structure), even as a morphological word projecting two syntactic nodes (courtesy of lexical sharing). Note also that the numeral itself can be complex. A different possibility is that Num N forms a productive, syntactic compound.

4.6.1 Welsh "compounds"

Welsh compounds can be subdivided into several types, depending on (apart from category of head): (i) whether they are head-initial or head-final (ii) whether they are loose or strict

1. In loose N N compounds, each member retains its own stress. If the first member is a FSG noun, the second member is soft-mutated.
2. Strict compounds have a single main stress, which is on the penultimate syllable, irrespective of the stress of the compound members when they appear independently:
   tywod 'sand' + *maen* 'stone' ⇒ *tywódfaen* 'sandstone'
   Word-internal phonotactic constraints apply (detail suppressed).

All combinations of the two parameters (strict/loose, head-initial/-final) exist, but:

[i]	I'll	←	D (⇓ SUBJ) =$_c$ ↓ (↓PRED) = PRO	I ⇓ = ↓ (↓TENSE) = FUT

- the only productive types are head-initial loose and head-final strict compounds
- head-final loose compounds are very rare (*uffern dân* 'hell fire' (usually used as a swear word))
- head-initial strict compounds are generally semantically opaque and clearly lexicalized

Head-initial loose compounds can be hard to distinguish from possessive construction with an indefinite possessor. The diagnostics that clearly identify a compound are: a) soft-mutation following a FSG noun, and b) appearance of a determiner in front of the entire compound.

Head-initial loose compounds freely admit coordinated nouns as members, unlike strict compounds.

(40) a. y Gwasanaeth Tân ac Achub
the service fire and rescue
'the Fire and Rescue Service'

b. y Fframwaith Gwasanaeth a Chyllid
the framework service and finance
'the Service and Financial Framework'

This is not possible with strict compounds. Where two coordinated strict compounds have a common member, it cannot be deleted under identity:

(41) *calch__ a thywod.faen
lime__ and sand.stone

(42) calch.faen a thywod.faen *(OK)*
lime.stone and sand.stone

4.6.2 Constructional Idioms

A possible treatment of Numeral Noun combinations might be as contructional idioms (Booij, 2009, 2002).

For e.g. Germanic complex predicates Booij proposes a syntactic template of the form [laten V] with two terminal positions one of which is fixed and the other is a variable. The advantage of this is that it can accommodate noncompositional meanings. This is a productive lexical template in the lexicon.

Some kinds of constructional idioms (such as those he hypothesises are the basis of productive, compositional A N compounds in Dutch) are functionally similar to compounds: "they form alternatives to morphological expression of lexical and/or grammatical content" (Booij, 2002, 312) and in fact feed word formation.[4] This class is an open set for which he proposes a treatment involving constructional idiom or template with the form $[\text{A} \quad \text{N}]_{NP}$ in the lexicon. Though it is not completely clear the idea is that this is a specific instantiation of a general syntactic template which inherits information (properties) from the general NP template.

So the idea would be that in fact there is no general PS rule introducing a prenominal Num at all. Rather there is a constructional idiom. However it will then be necessary to ensure that these are lexically instantiated nodes. Perhaps we can use the lexical instantiation function for this — the intended meaning is that these are nodes which immediately dominate lexical exponents.

N	→	Num	N
		(↑ NUMB) = ↓	↑ = ↓
		⇓	⇓

NB: recursion under Num to accomodate number strings: complex predicates?

[4] In Dutch you can distinguish AN compounds from A+N phrases because of adjective inflection. Of the ones which are not compounds, you can tell when they have a naming/classificatory function - it always involves a bare Adj.

5 Additional Complications?

(43) *pum llyfr a ffilm
five book and film

(44) hoff lyfrau a ffilmiau
favourite book and film

(45) pum hoff lyfr a ffilm
five favourite book and film

It may be outside the constructional idiom suggesting that the constructional idiom wins out only in potential Num N strings.

One speaker among the respondents appears to systematically allow distributive Num in the prenominal position. Perhaps such speakers have permitted the Num to occur more freely in the syntax, that is, beyond the constructional idiom, and treat it as a distributive feature.

References

Asudeh, Ash. 2002. The Syntax of Preverbal Particles and Adjunction in Irish. In M. Butt and T. H. King, eds., *The Proceedings of the LFG '02 Conference*. http://csli-publications.stanford.edu/LFG/7/lfg02asudeh.pdf.

Booij, Geert. 2002. Constructional idioms, morphology and the dutch lexicon. *Journal of Germanic Linguistics* 14:301–327.

Booij, Geert. 2009. Constructions and lexical units: an analysis of dutch numerals. *Linguistische Bericht* 19:1–14.

Bresnan, Joan and Sam A. Mchombo. 1995. The Lexical Integrity Principle: Evidence from Bantu. *Natural Language and Linguistic Theory* 13(2):181–254.

Dalrymple, Mary and Irina Nikolaeva. 2006. Syntax of natural and accidental coordination. *Language* 82-4:824–849.

Miller, Philip. 1992. Postlexical cliticization vs. affixation: coordination criteria. In G. C. C. Canakis and J. Denton, eds., *Papers from the 30th Chicago Linguistic Society*, pages 382–96. CLS.

Mittendorf, Ingo and Louisa Sadler. 2005. Nouns, Numerals and Number in Welsh NPs. In M. Butt and T. H. King, eds., *Proceedings of the LFG05 Conference*. Stanford, CA: CSLI Publications.

Thomas, Peter Wynn. 1996. *Gramadeg y Gymraeg*. Caerdydd: Gwasg Prifysgol Cymru.

Toivonen, Ida. 2003. *Non-Projecting Words: A Case Study of Swedish Particles*. Dordrecht: Kluwer Academic Publishers.

Wescoat, Michale. 2000. Udi Person Markers and Lexical Integrity. In M. Butt and T. H. King, eds., *On-line Proceedings of the LFG2009 Conference*.