The representation of group denoting nouns in a lexical knowledge base

Ann Copestake

University of Cambridge Computer Laboratory
New Museums Site, Pembroke Street, Cambridge, CB2 3QG, UK

January 1992 (revised version December 1992)

To appear in: P. St. Dizier and E. Viegas (editors) *Computational Lexical Semantics*, Cambridge University Press (in press)

Revised version of ESPRIT BRA-3030 ACQUILEX Working Paper No. 037

The representation of group denoting nouns in a lexical knowledge base

Ann Copestake

Abstract

In this paper we consider the representation of group denoting nouns, and the relationship between groups and their individual members, within the general framework of a computational treatment of lexical semantics which uses a typed unification based formalism to create a highly structured lexicon. We illustrate how such a representation can be created (semi-)automatically using machine readable dictionaries as a data source.

1 Introduction

The work reported here is part of research on the ACQUILEX project[1] which is aimed at the eventual development of a theoretically-motivated, but comprehensive and computationally tractable, multilingual lexical knowledge base (LKB) usable for natural language processing, lexicography and other applications. One of the goals of the ACQUILEX project was to demonstrate the feasibility of building an LKB by acquiring a substantial portion of the information semi-automatically from machine readable dictionaries (MRDs). We have paid particular attention to lexical semantic information. Our work therefore attempts to integrate several strands of research:

- Linguistic theories of the lexicon and lexical semantics. In this paper we will concentrate on the lexical semantics of nominals where our treatment is broadly based on that of Pustejovsky (1991), and in particular on his concepts of the *generative lexicon* and of *qualia structure*.

- Knowledge representation techniques. The formal lexical representation language (LRL) used in the ACQUILEX LKB system is based on typed features structures similar to those of Carpenter (1990, 1992), augmented with default inheritance and lexical rules. Our lexicons can thus be highly structured, hierarchical and generative.

- Lexicography and computational lexicography. The work reported here makes extensive use of the Longman Dictionary of Contemporary English (LDOCE; Procter, 1978). MRDs do not just provide data about individual lexical items; our theories of the lexicon have been developed and refined by considering the implicit organisation of dictionaries and the insights of lexicographers.

In this paper we will show how these strands can be combined in developing an appropriate representation for group nouns in the LRL, and in extracting the requisite information automatically from MRDs.

1.1 Qualia structure

Our treatment of the lexical semantics of nominals is ultimately based on Pustejovsky's approach, and, in particular, his description of *qualia structure*. Pustejovsky (1991) describes the qualia structure of a lexical item as consisting of the following four roles:

[1]ACQUILEX: 'The Acquisition of Lexical Knowledge for Natural Language Processing Systems': Esprit BRA-3030. For an overview of ACQUILEX as a whole see Briscoe (1991).

Constitutive Role The relation between an object and its constituents or proper parts.

Formal Role That which distinguishes the object within a larger domain.

Telic Role Purpose and function of the object.

Agentive Role Factors involved in the origin or "bringing about" of an object.

Pustejovsky argues that, rather than assuming that nominals just behave as passive objects when they combine with verbs, we should treat them as being as active in the semantics as the verb itself is. He refers to this behaviour as *cocompositionality*. The cocompositional behaviour of a noun is determined by its qualia structure.

For example, many verbs such as *enjoy* which can take a VP complement, can be described as selecting semantically for an event.

(1) Mary enjoys playing the guitar.

However *enjoy* can also take an NP complement, and in sentences such as (2) the complement *the book* apparently denotes an object.

(2) Mary enjoyed the book.

Traditionally the only way to handle this is to assume two lexical entries for *enjoy* and to relate the different senses by meaning postulates. However this is unsatisfactory since it leads to a proliferation of senses in the lexicon and it does not generalise to other cases where a noun phrase is interpreted as an event, such as (3).

(3) After three glasses of champagne, John felt much happier.

Furthermore, examples such as (4) seem perfectly acceptable:

(4) Mary enjoys books, television and playing the guitar.

However, under quite generally accepted assumptions about the nature of coordination and lexical ambiguity (e.g. Zwicky and Sadock, 1975; Cruse, 1986), only one sense of *enjoy* can be involved in (4).

Pustejovsky proposes that examples such as (2) be treated as involving *logical metonymy*. The sentence is interpreted as:

(5) Mary enjoyed some event associated with the book.

The qualia structure for nouns specifies possible associated events. In this case, for example, the telic (purpose) role of the qualia structure for *book* has a value equivalent to *reading*. When combined with *enjoy*, type coercion occurs, because *enjoy* selects for an event rather than an object, and the particular sort of event which is likely to be involved can be determined from the qualia structure, which results in a default interpretation for (2) equivalent to:

(6) Mary enjoyed reading the book.

In a marked context the default interpretation might be overridden or blocked. For example, if Mary was the name of a goat, it might be inferred that she enjoyed eating the book rather than reading it; this is assumed to be part of pragmatics rather than lexical semantics. In such cases the type coercion to an event still occurs but the nature of the event is overridden. In cases where the lexical semantics of the noun would not specify a telic role, or where the event specified would not be of an appropriate type, the corresponding sentences are odd:

(7) ? John enjoyed the rock.

Since *rock* is not regarded as having a specified telic role, the type of event which results from the metonymic coercion process is unknown, and the sentence seems bad.

For details of the way in which qualia structure and logical metonymy can be formally represented using a computationally tractable language, see Briscoe *et al.* (1990) and Copestake and Briscoe (1991). In the remainder of this paper, we will describe our current lexical representation language, and illustrate how we can use this to represent group nouns, in a way which integrates with the earlier accounts. As in the earlier papers, we concentrate on the issues of how the lexicon may be structured, in order to provide an efficient representation of lexical semantic information in the paradigmatic plane, and on how the description of lexical semantic structure may be integrated with a compositional semantic account. We conclude by describing how lexical information about group nouns may be extracted automatically from dictionary definitions.

2 The Lexical Representation Language

Our lexical representation language (LRL) is unification-based, allowing complex interconnections between syntactic and semantic information to be defined, and making a tight interface possible between the lexicon and other components of NLP systems, e.g. the parser/interpreter, the generator and even the transfer component in a machine translation system. In fact LRL is a slight misnomer, since grammar rules can also be written in the language and a parser is incorporated in the LKB system in order to test lexical entries. In designing the language we adopted a similar philosophy to that behind PATR-II (Shieber, 1986), in that the LRL is intended to be sufficiently general to encode a range of possible approaches to linguistic representation. Like PATR-II, the LRL is based on the use of feature structures. However, in contrast to PATR-II, feature structures are typed in a manner similar to that proposed by Carpenter (1990, 1992). Feature structures must be well-formed with respect to types and particular features will only be appropriate to specified types and their subtypes. The type system only allows non-default inheritance; we augment this with a restricted concept of default inheritance from feature structures, referred to as *psorts* in this context. Default inheritance is formalised in terms of default unification and is constrained by the type system. The LRL itself is a relatively 'theory neutral' language — a type system has to be developed to instantiate it in order to use it for representation based on a particular linguistic theory. The LRL is described in detail in various papers in Briscoe *et al.* (in press) and also in Copestake (1992). Here we will not attempt a formal description of the LRL but will give an informal overview, illustrated with relevant examples.[2] We will start by giving an example, which is intended to give an intuitive idea of the representation language and type system, and then consider the LRL in slightly more detail.

[2]The type system described in this paper was developed following on from that used on the ACQUILEX project as a whole, which was described in Copestake (1992), for example. The new system is somewhat simpler, it is less directly related to dictionary definitions and the treatment of some aspects of noun semantics has been improved.

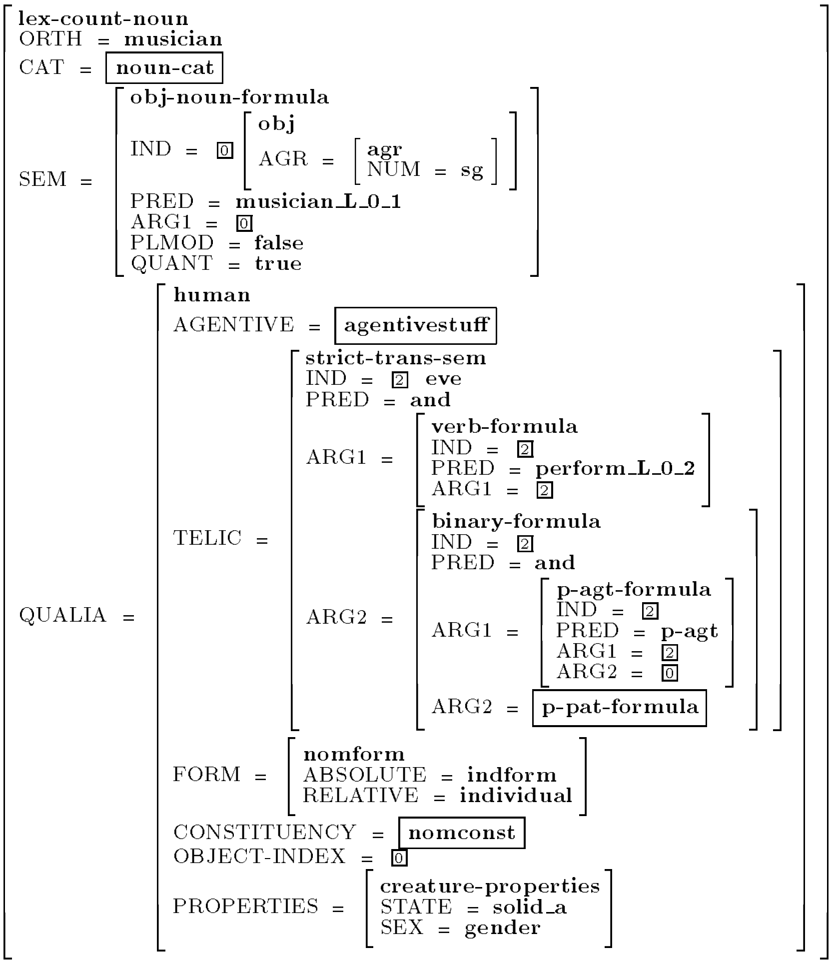

Figure 1: Lexical entry for musician

The following is a description of feature structure which corresponds to a lexical entry for *musician*:

musician L_0_1

```
< > < lex-individual < >
< QUALIA > < person_L_0_1 < QUALIA >
< QUALIA : TELIC  > <= perform_L_0_2 < SEM > .
```

The identifier, L_0_1, indicates that the entry is intended to (roughly) correspond to the LDOCE sense *musician 1*:

musician 1 a person who performs on a musical instrument ...

The feature structure into which this description expands is shown as an attribute-value matrix (AVM) in Figure 1.

The expansion is due to the combination of the type system, which defines the underlying 'templates' for all feature structures, and to (default) inheritance from psorts. Lower case bold font indicates types (e.g. **human**), upper case is used for features (e.g. QUALIA), in AVM diagrams, descriptions and text. Lower case typewriter font is used for psorts in descriptions and text (e.g. `person_L_0_1`, `lex-individual`). Reentrancy is indicated by a boxed integer in the AVM diagrams. Some parts of the entry are not shown fully expanded in this figure — this is indicated by a box round the type name.

$$\begin{bmatrix} \textbf{human} \\ \text{AGENTIVE} = \boxed{\textbf{agentivestuff}} \\ \text{TELIC} = \boxed{\textbf{verb-sem}} \\ \text{FORM} = \begin{bmatrix} \textbf{nomform} \\ \text{ABSOLUTE} = \textbf{indform} \\ \text{RELATIVE} = \textbf{(individual plural group)} \end{bmatrix} \\ \text{CONSTITUENCY} = \boxed{\textbf{nomconst}} \\ \text{OBJECT-INDEX} = \textbf{entity} \\ \text{PROPERTIES} = \begin{bmatrix} \textbf{creature-properties} \\ \text{STATE} = \textbf{solid\_a} \\ \text{SEX} = \textbf{gender} \end{bmatrix} \end{bmatrix}$$

Figure 2: Constraint on the lexical semantic type **human**

The description specified states that `musician_L_0_1` is to default inherit from the entire psort `lex-individual`, that its qualia structure is default inherited from the qualia structure of `person_L_0_1` and that the telic role is non-default inherited from the semantics of `perform_L_0_2`.

The lexical entry consists of four main components. The value for ORTH is a simple string representing the orthography. The syntactic component is indicated by the feature CAT, but is shown unexpanded here. We adopt a categorial approach to syntax, for details of which see Sanfilippo (in press). SEM introduces the formal semantic structure, which is encoded in a way which is basically equivalent to the lambda calculus expression $\lambda x[\textbf{musician\_L\_0\_1}(x)]$. Agreement is specified on the indices, following Pollard and Sag (forthcoming).

The feature QUALIA introduces the lexical semantic structure. The lexical semantic type of the entry is **human** (see Figure 2). The telic role is shown in detail (although it is not completely expanded); it has been instantiated with the semantics for the lexical entry for *perform*. Verb semantics are expressed in the type system as a whole in a neo-Davidsonian representation making use of thematic roles (see Sanfilippo, in press). The formula given for the telic role is equivalent to $\lambda e[\textbf{perform\_L\_0\_2}(e) \wedge \text{agent}(e, x)]$ where x is bound in the expression of the semantics of the lexical entry as a whole so that $\textbf{musician\_L\_0\_1}(x)$.

2.1 The type system

The type system is the basis for setting up linguistic representations in the LRL. The type hierarchy defines a partial ordering on the types and specifies which types are *consistent*. Only feature structures with mutually consistent types can be unified — two types which are unordered in the hierarchy are assumed to be inconsistent unless the user explicitly specifies a common subtype. Every *consistent* set of types has a unique greatest lower bound or meet; when two feature structures are successfully unified the type of the resulting feature structure will be the meet of their types. Thus, in the fragment of a type hierarchy shown in Figure 3, **natural** and **physical** are consistent; unifying a feature structure of type **natural** with one of type **physical** will result in a feature structure of type **natural_physical**.

Our formalism differs somewhat from that described by Carpenter in that we adopt a different notion of *well-formedness* of typed feature structures (i.e. consistency of feature structures with the type system). In the LRL, every type must have exactly one associated feature structure which acts as a constraint on all feature structures of that type; by subsuming all well-formed feature structures of that type. The constraint also defines

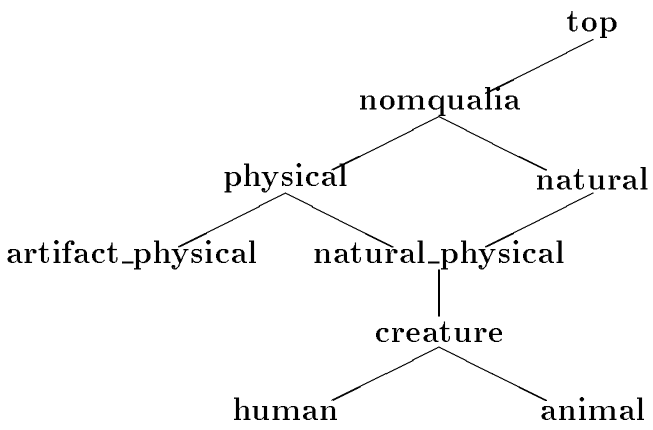

Figure 3: A fragment of the lexical semantic type hierarchy

which features are *appropriate* for a particular type — a well formed feature structure may only contain appropriate features. Constraints are inherited by all subtypes of a type, but a subtype may introduce new features (which will be inherited as appropriate features by all its subtypes). A constraint on a type is a well-formed feature structure of that type; all constraints must therefore be mutually consistent. The constraint on the type **human** is shown in Figure 2: in effect the information expressed in this is inherited non-defeasibly by the qualia structure of lexical entries such as that for *musician*. The type system has to be completely specified before any lexical entries can be expanded; this allows the well-formedness of lexical entries to be checked but is too inflexible to be the sole means of inheritance.

2.2 Psort inheritance

To allow default inheritance and more flexible non-default inheritance, we introduce the concept of a *psort*; a feature structure from which another feature structure inherits information, normally by default. The hierarchical ordering on psorts (which must be consistent with the type hierarchy) provides an order on defaults. Default inheritance is implemented by a version of default unification (e.g. Carpenter, in press). Multiple inheritance is restricted to the case where information inherited from different sources is consistent. Non-default inheritance from psorts is also allowed; this is simply implemented using ordinary unification. Default inheritance from a psort is indicated in the description language by `<`, non-default inheritance by `<=`.

Psorts may correspond to lexical entries or be specially defined in order to conveniently group some information. In the description for *musician* shown above `lex-individual`, `person_L_0_1` and `perform_L_0_2` are all psorts; the first is specially defined, the latter two are lexical entries. Because of the condition that the type hierarchy and the default inheritance hierarchy must be consistent, the default inheritance specification also determines the type of the qualia structure for *musician* to be **human**, non-defeasibly. The non-default inheritance from `perform_L_0_2` allows the appropriate semantic structure to be copied to fill the TELIC role. (Details of these psorts are shown in the appendix.) Other lexical entries, for example that for *minstrel*, will themselves inherit information taxonomically from the psort `musician_L_0_1` which is set up by the lexical entry. In the remainder of this paper, we will consider how group nouns may be represented in the LRL in a manner which is consistent with the rest of the type system.

3 Group nouns

Group nouns, such as *band*, *crowd*, *quartet*, *flock*, *management* and *group* itself, are distinctive in English in that, when morphologically singular, they behave in some respects like singular nouns and in others like plurals. This manifests itself in several ways:

1. Singular or plural pronouns can be used:

   (8) The band played well tonight. Its/their tour has sold out.

2. Either singular or plural agreement with the verb is possible (plural agreement with group nouns is, in general, less common in American English):

   (9) That band play/plays well.

   In the case of group nouns which denote groups of humans, a relative clause is introduced by *who* if plural agreement is used, and by *which* if it is not:

   (10) The band who get/*gets top billing at the festival receive/*receives £20,000.
   The band which gets/*get top billing at the festival receives/*receive £20,000.

3. Individual members can be referred to by using *one of* etc.

   (11) One of the band smashed her guitar.

The final criterion distinguishes between group nouns and those such as *barracks* and *gallows* which can take either singular or plural agreement (when referring to the same entity). Note also that unpluralised group nouns always take a singular determiner, even if verbal agreement is plural.

(12) This barracks is/*are new.
These barracks are/*is new.
That band has/have been playing well.

However, there are other nouns which do not meet these criteria, even though they refer to entities which can be regarded as being made up of several discrete individuals. For example, consider the LDOCE definition of *dolmen*:

dolmen a group of upright stones supporting a large flat piece of stone, built in ancient times in Britain and France

Despite the fact that a *dolmen* can evidently be regarded as a group of entities, it does not behave as a group noun; the following are all unacceptable:

(13) a The dolmen is on a mountain. *They're very eroded.
b *The dolmen have fallen down.
c *One of the dolmen fell down.

There is clearly a semantic distinction between group and non-group nouns; when a group noun is used the individual components of the entity denoted are sufficiently obvious that it can be referred to as though it were a plural term. Collectives such as *terrace* and *range*, which denote groups of entities of a particular type and which usually appear with *of* phrases (*terrace of houses*, *range of mountains*), do share some of the behaviour of group nouns, however. These nouns always take singular agreement when morphologically singular (at least when the *of* phrase is absent), and thus are not group nouns by the first two tests, but they can meet the third, although only in contexts where the individual members are explicitly mentioned.

(14) a The house was one of a terrace.
b * One of the terrace had a green front door.

Example (14a) is taken from the Lancaster-Oslo/Bergen (LOB) corpus; when checking for *one of* followed by a morphologically singular noun phrase, this was the only example where the head of the NP was not a group noun by the agreement tests. There is a contrast between *terrace* and *group*, as the latter is not at all limited in the semantic type of its *of* complement (e.g. *group of houses*, *group of statistics*, *group of actions*), but which refers to people when used without the *of* phrase in an unmarked context.

The singular/plural dual behaviour of true group nouns to some extent corresponds to whether the predicate is seen as applying to the group as a whole or to its individual members.

(15) The band was formed in 1977.
The team were killed in a plane crash.

There is a tendency for singular agreement to be used when the group as an entity is referred to, and for plural agreement to be used when the individuals are concerned. The examples above are odd when the agreement is changed:

(16) ? The band were formed in 1977.
? The team was killed in a plane crash.

Sometimes differences in agreement alone suggest a semantic distinction. In (17a) the implication is that the committee as an entity gets the money, (17b) suggests that it goes to the individual members and there is a possible distributive reading, forced in (17c). Note that (17d) is bad.

(17) a The committee gets £20,000 per annum.
b The committee get £20,000 per annum.
c The committee get £20,000 per annum each.
d ? The committee gets £20,000 per annum each.

However, plural agreement with verb phrases which apparently refer to the group as a single entity is quite normal in some contexts, such as when referring to sports teams or clubs.

(18) a Forfar are a good side. (LOB corpus)
b But there was to be no bargaining [on players' contracts]
as far as the club were concerned. (*The Guardian*)

It is useful to distinguish between ordinary group nouns and those which refer generically, since the latter class involve some different problems which we will not discuss here. Ordinary group nouns form plurals in the normal way, (e.g. *bands*, *crowds*, *quartets*, *flocks*), but others do not normally form plurals because they refer generically (e.g. *aristocracy*, *clergy*), or to an entity usually regarded as unique (e.g. *admiralty*), although plurals are possible in phrases such as *the admiralties of England and France*. The generic group nouns have dual group-entity/plural behaviour, but their plural behaviour parallels that of bare plural noun phrases in 'universal' position, in contrast to normal group nouns. For example (19a) implies (19b) but (19c) does not imply (19d), but only (19e).

(19) a The clergy are badly paid.
b Clergymen are badly paid.
c The committee are badly paid.
d Committeemen are badly paid.
e The committeemen are badly paid.

In what follows we will make the simplest assumption about the nature of the plural reading of ordinary group nouns, which is that it is equivalent to a normal plural. The plural reading corresponds to an entity which can be regarded as the sum of the members of the group, and has a qualia structure appropriate for a normal plural entity. [3] This straightforwardly accounts for the verbal and pronominal agreement, and for the use of partitives with the plural interpretation. On this assumption, we should get both distributive and collective plural readings. Distinguishing the group reading and the collective plural reading is not easy but there are examples such as (20a) which should probably be treated as a collective plural, since the predicate refers to individual members rather than the group entity. Similarly (20b) has a cumulative reading (Scha 1983), where the committee members are distributed in some unspecified way between the cars.

(20) a The committee are arriving in a car.
b The committee are arriving in three cars.

We then have to address the question of relating the group and the plural readings. We will assume that a group and the plural sum of its members are distinct entities, and that the plural reading is produced from the group reading by a process of logical metonymy, similar to that discussed in section 1.1. In this case, the instantiation involves the composition of the entity rather than its purpose. The metonymic account allows for the examples of group nouns such as *club*, *committee* and *company*, where the entity denoted seems to have an existence independent of its members, even when a purely extensional viewpoint is taken. One can imagine a club, for example, which currently has no members but nonetheless still exists as a legal entity. This is problematic for theories which make the representation the group entity dependent on its members, but does not pose any problems for the metonymic account. [4]

---

[3] In the formal semantics we treat the domain of individuals as having a lattice structure (Link, 1983). The semantics assumed are based on work by Krifka (1987). Plural individuals consist of a sum of ordinary individuals, but there is no distinction in formal semantic type between an ordinary individual and a plural one. Ordinary singular count predicates, such as **musician′**, denote sets of non-plural individuals. Plural predicates are formed by taking the closure of the denotation of the singular predicate, e.g. ***musician′**.

[4] The metonymic treatment is less plausible with group nouns such as *crowd* which cannot exist without members, and thus where the distinction between the group and its members has less justification on a purely extensional treatment. Landman (1989) has an extensive discussion of groups in the context of the treatment of plurality. For example, Landman treats the collective reading of sentences such as (21) as involving a group, rather than the plural sum.

So this suggests that in order to provide an adequate lexical semantics for group nouns, information about their membership must be represented, in order to allow appropriate semantic information to be associated with phrases such as *one of the band*. We associate this information with the CONSTITUENCY role of the qualia structure.

It appears that group nouns in English always refer to entities whose individual members are seen as capable of independent, agentive action, usually humans or other 'higher' animals. We have tested this assumption in a preliminary way using LDOCE; group nouns can, in theory, be retrieved as a class quite simply, since they are marked in the dictionary by the grammar codes GC (group countable, e.g. *committee*) or GU (group uncountable, e.g. *Admiralty*). Unfortunately the LDOCE coding is far from comprehensive in this case (*army*, *assembly*, *band*, *coven* have no senses marked as being group nouns, for example) and the GU code has been given to a considerable number of entries which would not be characterised as group nouns by the tests given above, especially plural forms such as *letters* and *tactics*. We thus considered only nouns with grammar code GC, and excluded the morphologically plural forms *games*, *Olympic Games* and *vibes*, which also do not meet all the tests. The remaining senses all refer to collections of humans or human organisations, or (less frequently) animals, with the exceptions *fleet* and *convoy*, where the individual entities are ships.[5] There are some group nouns which can refer to collections of people or organisations which may themselves be groups;

league[2] **3** a group of sports clubs or players ...

Thus grouping is not restricted to a single level.

It thus seems that lexicalisation of a concept of a collection of entities as a group nouns is restricted to a small semantic class of entities, which we will provisionally limit to humans, organisations and animals, ignoring the ship examples for the time being, since further work is necessary to more precisely delimit the class.

3.1 Representing group denoting nouns in the LRL

We can describe entries for group denoting nouns in the LRL, which allow us to formalise most of the aspects of their behaviour discussed above. An entry for *band*, in the sense meaning a group of musicians is shown in Figure 4. This feature structure corresponds to the group entity reading for *band*. Here we have assumed that number agreement is tightly linked to the group/plural distinction, and thus agreement is specified as **sg**. Alternatively it could be underspecified as **num** to allow for examples, such as those given earlier, where plural agreement is used in sentences which seem on semantic grounds to involve an uncoerced group entity.[6]

---

(21) John and Mary lifted a piano.

Landman makes a type distinction between a group and the plural sum of its members. However, since group formation can iterate indefinitely, this leads to a proliferation of types. Since we do not think that Landman's treatment adequately accounts for the properties of group nouns such as *club*, we have not adopted it here.

[5] There may be a connection here with the use of feminine gender personal pronouns when referring to ships, given that group nouns are normally associated with humans and higher animals.

[6] Neither of these options is completely satisfactory, of course. We would like to say that, by default, agreement is determined by the group/plural distinction, but that this default may be overridden. This is not possible in the current LRL, since defaults are part of the description language and thus operate only on the paradigmatic plane and do not affect syntagmatic combination.

$$
\left[\begin{array}{l}
\textbf{lex-count-noun} \\
\text{ORTH} = \textbf{band} \\
\text{CAT} = \boxed{\textbf{noun-cat}} \\
\text{SEM} = \left[\begin{array}{l}
\textbf{obj-noun-formula} \\
\text{IND} = \boxed{0} \left[\begin{array}{l} \textbf{obj} \\ \text{AGR} = \left[\begin{array}{l} \textbf{agr} \\ \text{NUM} = \textbf{sg} \end{array}\right] \end{array}\right] \\
\text{PRED} = \textbf{band\_L\_3\_2} \\
\text{ARG1} = \boxed{0} \\
\text{PLMOD} = \textbf{false} \\
\text{QUANT} = \textbf{true}
\end{array}\right] \\
\text{QUALIA} = \left[\begin{array}{l}
\textbf{human} \\
\text{AGENTIVE} = \boxed{\textbf{agentivestuff}} \\
\text{TELIC} = \left[\begin{array}{l}
\textbf{strict-trans-sem} \\
\text{IND} = \boxed{2}\ \textbf{eve} \\
\text{PRED} = \textbf{and} \\
\text{ARG1} = \left[\begin{array}{l} \textbf{verb-formula} \\ \text{IND} = \boxed{2} \\ \text{PRED} = \textbf{perform\_L\_0\_2} \\ \text{ARG1} = \boxed{2} \end{array}\right] \\
\text{ARG2} = \boxed{\textbf{binary-formula}}
\end{array}\right] \\
\text{FORM} = \left[\begin{array}{l} \textbf{nomform} \\ \text{RELATIVE} = \textbf{group} \end{array}\right] \\
\text{CONSTITUENCY} = \left[\begin{array}{l} \textbf{groupconst} \\ \text{ELEMENTS} = \boxed{\textbf{human}} \end{array}\right] \\
\text{OBJECT-INDEX} = \boxed{0}
\end{array}\right]
\end{array}\right]
$$

Figure 4: Lexical entry for *band*

The lexical semantic type of the entry is **human**, but the values for the features CONSTITUENCY and FORM : RELATIVE are specific to group denoting nouns. The representation for the lexical semantics of group denoting nouns has to be compatible with the rest of the type system. Given the results above, which suggest that there are only restricted semantic classes of group nouns, it clearly would be inappropriate to parallel the entire existing lexical semantic type hierarchy with a group type hierarchy. Furthermore much of the information about group nouns will be comparable with that about their individual members. We therefore allow types such as **human** to apply to both individuals and groups, and distinguish between the two by specifying that the CONSTITUENCY feature either takes type **nongroupconst** or **groupconst**. Only the latter has ELEMENTS as an appropriate feature. The lexical semantics of the plural sum of the individuals making up a group noun is specified as the value of the ELEMENTS feature. In the case of *band* the individuals involved are *musicians*, thus the ELEMENTS slot is instantiated with the qualia structure corresponding to the pluralised form of `musician_L_0_1`. (In Figure 4 this is shown unexpanded so only the type, **human**, is apparent.) The value of FORM : RELATIVE is also specific to group denoting nouns. In general FORM : RELATIVE specifies individuation relative to a predicate — other possible values include **individual**, **plural** and **mass**.

The lexical entry given above does not directly account for the plural reading of group nouns. Our treatment of the group/members logical metonymy is very similar to that of the entity/event coercion, but in this case type coercion will occur in contexts where a plural entity is required. We implement this in LAUREL with a unary rule, `group-to-plural`, which applies to a group denoting noun phrase. Rules in the LRL are themselves feature structures, which can be taken as describing the relationship between an input structure and an output structure. The rule `group-to-plural` is actually quite complex, since it has to apply to type-raised noun phrases. It is given in full in the appendix — here we will summarise its effects on the various parts of the sign:

$$
\left[\begin{array}{l}
\textbf{nominal-sign} \\
\text{ORTH} = \left[\begin{array}{l}\textbf{complex-orth}\\ \text{ORTH1} = \textbf{the}\\ \text{ORTH2} = \textbf{band}\end{array}\right] \\
\text{CAT} = \boxed{\textbf{raised-np-cat}} \\
\text{SEM} = \left[\begin{array}{l}
\textbf{binary-formula}\\
\text{IND} = \boxed{0}\ \textbf{entity}\\
\text{PRED} = \textbf{the}\\
\text{ARG1} = \left[\begin{array}{l}
\textbf{binary-formula}\\
\text{IND} = \boxed{1}\left[\begin{array}{l}\textbf{dummy-or-obj}\\ \text{AGR} = \left[\begin{array}{l}\textbf{agr}\\ \text{NUM} = \textbf{pl}\end{array}\right]\end{array}\right]\\
\text{PRED} = \textbf{the\_1}\\
\text{ARG1} = \left[\begin{array}{l}
\textbf{obj-noun-formula}\\
\text{IND} = \boxed{2}\left[\begin{array}{l}\textbf{obj}\\ \text{AGR} = \left[\begin{array}{l}\textbf{agr}\\ \text{NUM} = \text{sg}\end{array}\right]\end{array}\right]\\
\text{PRED} = \textbf{band\_L\_3\_2}\\
\text{ARG1} = \boxed{2}\\
\text{PLMOD} = \textbf{false}\\
\text{QUANT} = \textbf{true}
\end{array}\right]\\
\text{ARG2} = \left[\begin{array}{l}
\textbf{binary-formula}\\
\text{IND} = \boxed{1}\\
\text{PRED} = \textbf{membership}\\
\text{ARG1} = \boxed{1}\\
\text{ARG2} = \boxed{2}
\end{array}\right]
\end{array}\right]\\
\text{ARG2} = \boxed{3}\left[\begin{array}{l}
\textbf{formula}\\
\text{IND} = \boxed{0}\\
\text{PRED} = \textbf{logical-pred}\\
\text{ARG1} = \textbf{sem}
\end{array}\right]
\end{array}\right] \\
\text{QUALIA} = \boxed{\textbf{human}}
\end{array}\right]
$$

Figure 5: Coerced form of *the band*

- The orthography is unchanged.
- The categorial syntactic structure remains of type **raised-np-cat**, but is changed so that it integrates appropriately with the new semantics and qualia structure.
- The formal semantics is set up so that it is essentially equivalent to 'the members of [input NP]' (e.g. *the members of the band*). The operator **membership** applies to the group entity, to give the plural entity.
- The output sign has plural agreement.
- The qualia structure for the output sign is equal to the value of `< QUALIA : CONSTITUENCY : ELEMENTS >` in the input sign. This will have the FORM appropriate for a plural entity.

The sign which would result from the application of `group-to-plural` to *the band* is shown in Figure 5. This sign has obligatory plural agreement and denotes the plural entity which consists of the members of the band. The unary rule application is forced in contexts where a formally plural entity is required, for example a partitive construction, such as *one of*. When this is used with a group noun (e.g. *one of the band*) the unary rule is applied to the group to give a plural entity, and the appropriate specification of the individuals involved is then produced in much the same way as for the ordinary plural (e.g. *one of the members of the band*).

The representation of lexical semantic information about the individuals which comprise the group thus allows the plural-like aspects of the behaviour of group nouns to be accounted for. The `group-to-plural` rule specifies the qualia structure of the resulting plural entity according to the composition described in the CONSTITUENCY feature and the effect on the lexical semantics thus parallels the effect on the logical representation.

4 The use of MRDs

To some extent, lexical entries such as those shown in this paper can be acquired semi-automatically from LDOCE. To do this we make use of a combination of information from the definition and from the LDOCE grammar codes. For example, consider again the definition of *musician* and the corresponding LRL entry:

musician 1 a person who performs on a musical instrument ...

```
musician L_0_1
   < > < lex-individual < >
   < QUALIA > < person_L_0_1 < QUALIA >
   < QUALIA : TELIC  > <= perform_L_0_2 < SEM > .
```

Syntactic information can be derived from LDOCE's grammar codes; in this case *musician* corresponds to an ordinary count noun, which results in the inheritance from `lex-individual`. We will not discuss extraction of information from grammar codes in detail here.

Noun definitions can be split into a *genus* and *differentia*; the genus can usually be taken to be the syntactic head of the definition. Here the genus term is taken to be (a particular sense of) *person*. The basis for our use of noun dictionary definitions as a source of lexical semantic information is that, in general, we can specify that entries inherit lexical semantic information by default from the entry corresponding to their genus term. The differentia may augment or override the information inherited from the genus term.

We can automatically extract genus terms reliably from noun definitions such as that above (Vossen, 1990) and lexically disambiguate them (semi-)automatically (Copestake, 1990).[7] Extracting information from the differentia is more difficult. It is possible, once the semantic type of a sense is known, to use the type as a template to guide analysis of the definitions, but a considerable amount of work is required to achieve this, and in ACQUILEX so far this has only been attempted on limited classes of definitions (see, for example, Ageno *et al.*, 1992; Vossen, 1992). Thus the telic role for the entry above was manually specified. In general, associating information manually with lexical entries which are frequently used as psorts is an effective way of acquiring information, since inheritance will result in a large number of other entries being instantiated. However there are many special cases where definitions do not straightforwardly yield genus terms which can be interpreted as psorts. This is extensively discussed in Vossen and Copestake (in press); here I will concentrate on the particular case of group nouns.

4.1 Dictionary definitions of group nouns

It is usually assumed that dictionary definitions should, in general, be substitutable, in context, for the word being defined (e.g. Landau, 1984). Because of this principle of

[7]The disambiguation procedure is semi-automatic in that a series of heuristics are used to determine the sense of the genus term, and the user is asked to confirm the choice made by the heuristics in some cases, to avoid large numbers of lexical entries inheriting information incorrectly. When creating hierarchies of concrete nouns from LDOCE, the user checks about 5% of the entries.

substitutability, definitions of group nouns in dictionaries such as LDOCE will normally be group denoting noun phrases. In some cases the genus term will be a relatively specific group noun, for example:

crew ^{1} **3** a rowing team

Such definitions can be treated as illustrated above: the entire qualia structure is default inherited from the entry for the genus sense, *team 2*.

crew L_1_3

```
< > < lex-group < >
 < QUALIA > < team_L_0_2 < QUALIA >.
```

However there is another class of definitions where the genus phrase is of the form 'DET *group of* N ' and the noun is principally being defined in terms of its members, for example:

band^{3} **2** a group of musicians ...

Such cases pose more problems, since there is very little semantic information that can be inherited from *group*. As illustrated with the example of *dolmen*, given earlier, the use of *group of* does not necessarily indicate a noun which is group denoting in the technical sense. In other cases, a group noun may be defined using a plural genus term, for example:

audience 1 the people listening to or watching a performance, speech, television show, etc.

In this case, the definition cannot be substituted for the *audience* in contexts where it is used with singular agreement or refers to the group as a whole.

(22) The audience were very noisy tonight.
The people listening to the performance were very noisy tonight.
The audience was very noisy tonight.
*The people listening to the performance was very noisy tonight.
The audience was tiny.
*The people listening to the performance was tiny.

(Presumably the lexicographer felt that it was better to use *people* than *group of people*, for example, which perhaps suggests a greater cohesion between the individuals than is appropriate here.) For such examples, a representation has to be built based on information about the individual members, rather than about the group as a whole.

Clearly, given the type system outlined earlier, information about the type of a group's members makes it possible to infer the type of the group noun. If the members are of type **human** then the group as a whole will be of type **human**, and so on. Furthermore the CONSTITUENCY role can be instantiated with the qualia structure for the members. So the following entry could be produced for *band*^{3} *2*:

band L_3_2

```
< > < lex-group < >
< QUALIA > = human
< QUALIA : CONSITUENCY : ELEMENTS >
          < ( musician_L_0_1 + plural ) < QUALIA >.
```

Here (`musician_L_0_1 + plural`) indicates that the rule for plural formation is applied to the psort `musician_L_0_1` before inheritance of the qualia structure takes place (see appendix).

However this leaves some information unspecified, in particular the TELIC role. We assume that this can be inherited from the TELIC role of the members, so the TELIC role of *band*[3] *2* is inherited from *musician* for example,

```
< QUALIA : TELIC >
       < musician_L_0_1 < QUALIA : TELIC >.
```

Adding this to the description above we get a specification which will expand out into the feature structure shown in Figure 4.

The inheritance of the telic role needs some justification, since in general usage it is not possible to assume that a group as a whole has a property even if all its members have that property, and the property is one which could hold of the group as a whole. For example:

> All the members of the committee are against the poll tax.

does not entail that:

> The committee is against the poll tax.

Inheritance of the telic role might also be problematic. There could be a group of musicians who got together to play football, for example, so:

> The King's Road football team is a group of musicians.

could be true, in which case the purpose of the group described would not be equivalent to that lexically specified by *musician*. However such examples are exceptional, and in the special case of dictionary definitions, if a group is defined in terms of its members it can be taken to inherit appropriate properties from them; we have not found any counter-examples to this in LDOCE so far. Intuitively, it seems unlikely that a dictionary would define a group in terms of its members, if they had a purpose or function which was distinct from that of the group. Even if a concept such as the musicians' football team were lexicalised, the lexicographer would have to specify the function of the group explicitly to avoid being misleading, and thus the default inheritance from its members' telic role would be overridden.

Thus, on the assumption that we can recognise the class of group nouns, and distinguish their members in the definitions, we can (semi-)automatically extract reasonably adequate entries, such as that shown for *band*. This is quite straightforward for nouns which are specified as GC in the LDOCE grammar coding scheme and which have entries of the form, 'DET *group of* N '. However there are problems in recognising the class of group nouns. As we mentioned earlier, the grammar coding scheme has not been applied consistently to group nouns. There are a variety of ways in which group nouns can be defined, and some of these patterns of definitions can also be used for non-group nouns. Although we can use a range of heuristics to identify candidate group nouns, these rely on certain assumptions, for example that any noun which denotes more than one person, but which is not marked as plural, is likely to be a group noun. Clearly, we need other sources of information in order to attempt comprehensive extraction.

Other MRDs could be used to enhance our existing data, but since sense-to-sense mapping on independent dictionary sources is a hard problem, which itself requires some form of LKB, we could at best determine that a headword has some sense which is marked as being a group noun, and use this as an additional heuristic. Corpora could also be used, but a massive amount of data is needed to have a significant chance of finding the less frequent group nouns used in a context where their distinctive behaviour is apparent. There are about 30 occurrences of *crowd* as a singular noun in the approximately 1.2 million word LOB corpus, and only one of these is in a context where singular/plural agreement can be distinguished. Checking a range of nouns manually for dual agreement on corpora of sufficient size would be an extremely labour intensive task without tools to partially parse selected sentences. Even then, the problem of sense distinction still remains — the skills of the professional lexicographer are really needed here. We believe that the way forward for computational lexicology and lexicography is collaboration with lexicographers and dictionary publishers, giving them the tools with which to instantiate LKBs, which could then be used for linguistic research, for NLP, and for the production of conventional dictionaries.

5 Conclusion

We have used the example of group nouns in this paper to illustrate the way in which we attempt to combine ideas and techniques from lexical semantics, lexicography and knowledge representation. Although the fragment shown here is integrated into a system which treats other complex aspects of lexical semantics, such as logical metonymy and sense extension, much more work is clearly needed to produce a comprehensive treatment of noun semantics which would properly test our approach. Furthermore there are some issues which appear to require modifications to the LRL, in particular with respect to the treatment of defaults. We are currently investigating richer representational frameworks, which should, for example, allow a better treatment of agreement than that shown here. Briscoe *et al.* (this volume) discusses one way in which a more general notion of defaults can be combined with a feature structure based representation language.

6 References

Ageno, A., I. Castellon, G. Rigau, H. Rodriguez, M.F. Verdejo, M.A. Marti, M. Taule (1992) 'SEISD: an environment for extraction of semantic information from on-Line dictionaries', *Proceedings of the 3rd Conference on Applied Natural Language Processing (ANLP-92),* Trento, Italy, pp. 253–255.

Briscoe, E. J. (1991) 'Lexical issues in natural language processing' in E. Klein and F. Veltman (ed.), *Natural language and speech,* Springer-Verlag, pp. 39–68.

Briscoe, E. J., A. Copestake and B. Boguraev (1990) 'Enjoy the paper: lexical semantics via lexicology', *Proceedings of the 13th International Conference on Computational Linguistics (COLING-90),* Helsinki, pp. 42–47.

Briscoe, E. J., A. Copestake and V. de Paiva (eds) (in press) *Inheritance, defaults and the lexicon,* Cambridge University Press.

Carpenter, R. (1990) 'Typed feature structures: inheritance, (in)equality and extensionality', *Proceedings of the First International Workshop on Inheritance in Natural Language Processing,* Tilburg, The Netherlands, pp. 9–18.

Carpenter, R. (1992) *The logic of typed feature structures,* Cambridge University Press.

Carpenter, R. (in press) 'Skeptical and credulous default unification with applications to templates and inheritance' in E. J. Briscoe, A. Copestake and V. de Paiva (ed.), *Inheritance, defaults and the lexicon,* Cambridge University Press.

Copestake, A. (1990) 'An approach to building the hierarchical element of a lexical knowledge base from a machine readable dictionary', *Proceedings of the First International Workshop on Inheritance in Natural Language Processing,* Tilburg, The Netherlands, pp. 19–29.

Copestake, A. (1992) 'The ACQUILEX LKB: representation issues in semi-automatic acquisition of large lexicons', *Proceedings of the 3rd Conference on Applied Natural Language Processing (ANLP-92),* Trento, Italy, pp. 88–96.

Copestake, A. and E. J. Briscoe (1991) 'Lexical operations in a unification based framework' in J. Pustejovsky and S. Bergler (ed.), *Lexical Semantics and Knowledge Representation. Proceedings of the first SIGLEX Workshop, Berkeley, CA,* Springer-Verlag, Berlin, pp. 101–119.

Cruse, D. A. (1986) *Lexical semantics,* Cambridge University Press, Cambridge, England.

Krifka, M. (1987) 'Nominal reference and temporal constitution: towards a semantics of quantity', *Proceedings of the 6th Amsterdam Colloquium,* University of Amsterdam, pp. 153–173.

Landau, S. I. (1984) *Dictionaries: the art and craft of lexicography,* Scribner, New York.

Landman, F. (1989) 'Groups I + II', *Linguistics and Philosophy,* **12 (5,6)**, 559–606, 723–744.

Link, G. (1983) 'The logical analysis of plurals and mass terms: a lattice-theoretical approach' in Bäuerle, Schwarze and von Stechow (ed.), *Meaning, use and interpretation of language,* de Gruyter, Berlin, pp. 302–323.

Procter, P. (editor) (1978) *Longman Dictionary of Contemporary English,* Longman, England.

Pustejovsky, J. (1991) 'The generative lexicon', *Computational Linguistics,* **17(4)**, 409–441.

Sanfilippo, A. (in press) 'LKB encoding of lexical knowledge from machine-readable dictionaries' in E. J. Briscoe, A. Copestake and V. de Paiva (ed.), *Inheritance, defaults and the lexicon,* Cambridge University Press.

Scha, R.J.H. (1983) 'Logical foundations for question answering', PhD thesis.

Shieber, S. M. (1986) *An introduction to unification-based approaches to grammar,* CSLI Lecture Notes 4, Stanford CA.

Vossen, P. (1990) 'A parser-grammar for the meaning descriptions of LDOCE', Links Project Technical Report 300-169-007, Amsterdam University.

Vossen, P. (1992) 'An empirical approach to automatically construct a knowledge base from dictionaries', *Proceedings of the 5th EURALEX,* Tampere, Finland.

Vossen, P. and A. Copestake (in press) 'Untangling definition structure into knowledge representation' in E. J. Briscoe, A. Copestake and V. de Paiva (ed.), *Inheritance, defaults and the lexicon,* Cambridge University Press.

Zwicky, A. M. and J. M. Sadock (1975) 'Ambiguity tests and how to fail them' in J. P. Kimball (ed.), *Syntax and Semantics IV,* Academic Press, New York, pp. 1–36.

Appendix

In this appendix we give more detailed descriptions of the structures used in the examples given in the paper.

Types, psorts and lexical entries

The following structure shows the type, **lex-count-noun**, which is the type of all the lexical entries described:

$$
\left[\begin{array}{l}
\textbf{lex-count-noun} \\
\text{ORTH} = \textbf{orth} \\
\text{CAT} = \boxed{\textbf{noun-cat}} \\
\text{SEM} = \left[\begin{array}{l}
\textbf{obj-noun-formula} \\
\text{IND} = \boxed{0} \left[\begin{array}{l} \textbf{obj} \\ \text{AGR} = \left[\begin{array}{l} \textbf{agr} \\ \text{NUM} = \textbf{number} \end{array}\right] \end{array}\right] \\
\text{PRED} = \boxed{1}\ \textbf{logical-pred} \\
\text{ARG1} = \boxed{0} \\
\text{PLMOD} = \textbf{boolean} \\
\text{QUANT} = \textbf{boolean}
\end{array}\right] \\
\text{QUALIA} = \left[\begin{array}{l}
\textbf{nomqualia} \\
\text{AGENTIVE} = \textbf{nomagent} \\
\text{TELIC} = \boxed{\textbf{verb-sem}} \\
\text{FORM} = \left[\begin{array}{l} \textbf{nomform} \\ \text{ABSOLUTE} = \textbf{real-form} \\ \text{RELATIVE} = \textbf{countable} \end{array}\right] \\
\text{CONSTITUENCY} = \boxed{\textbf{nomconst}} \\
\text{OBJECT-INDEX} = \boxed{0}
\end{array}\right]
\end{array}\right]
$$

For current purposes the important points about this type are:

1. A value of **true** for the feature PLMOD in the semantics is intended to be interpreted as equivalent to modifying the predicate by the closure operator *. The value **false** indicates that the predicate is not so modified.

2. The feature QUANT encodes the cumulative/quantised distinction (see Krifka 1987).

3. The type **countable** given as the value of the relative form in the qualia structure has several subtypes — those relevant here are **individual**, **plural** and **group**.

All noun lexical entries are specified as inheriting from one of a range of psorts which specify their mode of individuation. Those relevant here are `lex-individual`, `lex-plural` and `lex-group`:

lex-individual

`< > =` **lex-count-noun**
`< QUALIA : FORM : RELATIVE > =` **individual**
`< SEM : IND : AGR : NUM > =` sg
`< SEM : PLMOD > =` **false**
`< SEM : QUANT > =` **true** .

lex-plural

`< > =` **lex-count-noun**
`< QUALIA : FORM : RELATIVE > =` **plural**
`< SEM : QUANT > =` **false**
`< SEM : IND : AGR : NUM > =` **pl**.

lex-group

```
< > < lex-individual < >
< QUALIA : FORM : RELATIVE > = group
< QUALIA : CONSTITUENCY > = groupconst.
```

Since these are psorts, values may be overridden. This allows for nouns like *barracks* or *gallows* which are treated as basically individual denoting, although they may take plural agreement.

The relevant parts of the lexical entries for *person* and *perform* from which the entry for *musician* inherits are shown below.

$$\left[\begin{array}{l}
\textbf{lex-count-noun} \\
\text{ORTH} = \textbf{person} \\
\text{CAT} = \boxed{\textbf{noun-cat}} \\
\text{SEM} = \boxed{\textbf{obj-noun-formula}} \\
\text{QUALIA} = \left[\begin{array}{l}
\textbf{human} \\
\text{AGENTIVE} = \left[\begin{array}{l} \textbf{agentivestuff} \\ \text{ORIGIN} = \textbf{basic} \end{array}\right] \\
\text{TELIC} = \boxed{\textbf{verb-sem}} \\
\text{FORM} = \left[\begin{array}{l} \textbf{nomform} \\ \text{ABSOLUTE} = \textbf{real-form} \\ \text{RELATIVE} = \textbf{individual} \end{array}\right] \\
\text{CONSTITUENCY} = \boxed{\textbf{nomconst}} \\
\text{OBJECT-INDEX} = \boxed{1} \left[\begin{array}{l} \textbf{obj} \\ \text{AGR} = \left[\begin{array}{l} \textbf{agr} \\ \text{NUM} = \textbf{sg} \end{array}\right] \end{array}\right] \\
\text{PROPERTIES} = \left[\begin{array}{l} \textbf{creature-properties} \\ \text{STATE} = \textbf{solid_a} \\ \text{SEX} = \textbf{gender} \end{array}\right]
\end{array}\right]
\end{array}\right]$$

$$\left[\begin{array}{l}
\textbf{strict-trans-sign} \\
\text{ORTH} = \textbf{perform} \\
\text{CAT} = \boxed{\textbf{strict-trans-cat}} \\
\text{SEM} = \left[\begin{array}{l}
\textbf{strict-trans-sem} \\
\text{IND} = \boxed{0}\ \textbf{eve} \\
\text{PRED} = \textbf{and} \\
\text{ARG1} = \left[\begin{array}{l} \textbf{verb-formula} \\ \text{IND} = \boxed{0} \\ \text{PRED} = \boxed{1}\ \textbf{perform_L_0_2} \\ \text{ARG1} = \boxed{0} \end{array}\right] \\
\text{ARG2} = \left[\begin{array}{l}
\textbf{binary-formula} \\
\text{IND} = \boxed{0} \\
\text{PRED} = \textbf{and} \\
\text{ARG1} = \boxed{2} \left[\begin{array}{l} \textbf{p-agt-formula} \\ \text{IND} = \boxed{0} \\ \text{PRED} = \textbf{p-agt} \\ \text{ARG1} = \boxed{0} \\ \text{ARG2} = \left[\begin{array}{l} \textbf{obj} \\ \text{AGR} = \left[\begin{array}{l} \textbf{agr} \\ \text{NUM} = \textbf{number} \end{array}\right] \end{array}\right] \end{array}\right] \\
\text{ARG2} = \boxed{3} \left[\begin{array}{l} \textbf{p-pat-formula} \\ \text{IND} = \boxed{0} \\ \text{PRED} = \textbf{p-pat} \\ \text{ARG1} = \boxed{0} \\ \text{ARG2} = \left[\begin{array}{l} \textbf{obj} \\ \text{AGR} = \left[\begin{array}{l} \textbf{agr} \\ \text{NUM} = \textbf{number} \end{array}\right] \end{array}\right] \end{array}\right]
\end{array}\right]
\end{array}\right]
\end{array}\right]$$

Unary rules

The examples in the paper use two unary rules, `group-to-plural` and `plural`. Unary rules in the LRL are feature structures of type **unary-rule**. This has two features 1 and 0 which indicate the input and output of the rule respectively. Unary rules are used to encode type shifting, morphological rules and sense extensions — the only distinction

$$\left[\begin{array}{l}
\textbf{nominal-sign} \\
\text{ORTH} = \left[\begin{array}{l}\textbf{complex-orth}\\ \text{ORTH1} = \textbf{the}\\ \text{ORTH2} = \textbf{band}\end{array}\right] \\
\text{CAT} = \boxed{\textbf{raised-np-cat}} \\
\text{SEM} = \left[\begin{array}{l}
\textbf{binary-formula}\\
\text{IND} = \boxed{0}\ \textbf{entity}\\
\text{PRED} = \textbf{the_1}\\
\text{ARG1} = \left[\begin{array}{l}
\textbf{obj-noun-formula}\\
\text{IND} = \boxed{1}\left[\begin{array}{l}\textbf{obj}\\ \text{AGR} = \left[\begin{array}{l}\textbf{agr}\\ \text{NUM} = \textbf{sg}\end{array}\right]\end{array}\right]\\
\text{PRED} = \textbf{band_1_3_2}\\
\text{ARG1} = \boxed{1}\\
\text{PLMOD} = \textbf{false}\\
\text{QUANT} = \textbf{true}
\end{array}\right]\\
\text{ARG2} = \boxed{2}\left[\begin{array}{l}\textbf{formula}\\ \text{IND} = \boxed{0}\\ \text{PRED} = \textbf{logical-pred}\\ \text{ARG1} = \textbf{sem}\end{array}\right]
\end{array}\right] \\
\text{QUALIA} = \boxed{3}\left[\begin{array}{l}
\textbf{human}\\
\text{AGENTIVE} = \boxed{\textbf{agentivestuff}}\\
\text{TELIC} = \boxed{\textbf{strict-trans-sem}}\\
\text{FORM} = \left[\begin{array}{l}\textbf{nomform}\\ \text{ABSOLUTE} = \textbf{real-form}\\ \text{RELATIVE} = \textbf{group}\end{array}\right]\\
\text{CONSTITUENCY} = \left[\begin{array}{l}\textbf{groupconst}\\ \text{PARTICLES} = \textbf{string}\\ \text{ELEMENTS} = \boxed{\textbf{human}}\\ \text{UNITQUANTITY} = \textbf{string}\end{array}\right]\\
\text{OBJECT-INDEX} = \boxed{1}\\
\text{PROPERTIES} = \boxed{\textbf{creature-properties}}
\end{array}\right]
\end{array}\right]$$

Figure 6: AVM diagram for the phrasal sign *the band*

between these types of rule is that morphological rules are the only ones which involve orthographic changes and that type shifting rules may involve phrasal signs, whereas morphological and sense extension rules are limited to lexical signs. Thus `group-to-plural` is a type shifting rule which takes a phrasal sign as input (in this case an NP) but `plural` is a morphological rule. The sign for the unshifted NP, *the band*, is shown in Figure 6. It can be seen that this will unify with the the input section of the unary rule `group-to-plural` which is shown in Figure 7, giving the feature structure shown in Figure 5 as output.

The plural rule which was referred to in Section 4 is shown in Figure 8 for completeness.

$$
\left[\begin{array}{l}
\mathbf{unary\text{-}rule} \\
0 = \left[\begin{array}{l}
\mathbf{nominal\text{-}sign} \\
\text{ORTH} = \boxed{0}\ \mathbf{orth} \\
\text{CAT} = \left[\begin{array}{l}
\mathbf{raised\text{-}np\text{-}cat} \\
\text{RESULT} = \boxed{1}\ \boxed{\mathbf{sign}} \\
\text{DIRECTION} = \boxed{2}\ \mathbf{direction} \\
\text{ACTIVE} = \left[\begin{array}{l}
\mathbf{sign} \\
\text{ORTH} = \mathbf{orth} \\
\text{CAT} = \left[\begin{array}{l}
\mathbf{complex\text{-}cat} \\
\text{RESULT} = \boxed{1} \\
\text{DIRECTION} = \boxed{2} \\
\text{ACTIVE} = \boxed{\mathbf{nominal\text{-}sign}}
\end{array}\right] \\
\text{SEM} = \boxed{3}\ \boxed{\mathbf{formula}}
\end{array}\right]
\end{array}\right] \\
\text{SEM} = \left[\begin{array}{l}
\mathbf{binary\text{-}formula} \\
\text{IND} = \boxed{4}\ \mathbf{entity} \\
\text{PRED} = \mathbf{the} \\
\text{ARG1} = \boxed{5} \left[\begin{array}{l}
\mathbf{binary\text{-}formula} \\
\text{IND} = \boxed{6} \left[\begin{array}{l}
\mathbf{dummy\text{-}or\text{-}obj} \\
\text{AGR} = \left[\begin{array}{l}
\mathbf{agr} \\
\text{NUM} = \mathbf{pl}
\end{array}\right]
\end{array}\right] \\
\text{PRED} = \mathbf{logical\text{-}pred} \\
\text{ARG1} = \boxed{\mathbf{formula}} \\
\text{ARG2} = \left[\begin{array}{l}
\mathbf{binary\text{-}formula} \\
\text{IND} = \mathbf{entity} \\
\text{PRED} = \mathbf{membership} \\
\text{ARG1} = \boxed{6} \\
\text{ARG2} = \boxed{7}\ \mathbf{entity}
\end{array}\right]
\end{array}\right] \\
\text{ARG2} = \boxed{3}
\end{array}\right] \\
\text{QUALIA} = \boxed{8} \left[\begin{array}{l}
\mathbf{nomqualia} \\
\text{AGENTIVE} = \mathbf{nomagent} \\
\text{TELIC} = \boxed{\mathbf{verb\text{-}sem}} \\
\text{FORM} = \left[\begin{array}{l}
\mathbf{nomform} \\
\text{ABSOLUTE} = \mathbf{real\text{-}form} \\
\text{RELATIVE} = \mathbf{plural}
\end{array}\right] \\
\text{CONSTITUENCY} = \boxed{\mathbf{nomconst}} \\
\text{OBJECT-INDEX} = \mathbf{entity}
\end{array}\right]
\end{array}\right] \\
1 = \left[\begin{array}{l}
\mathbf{nominal\text{-}sign} \\
\text{ORTH} = \boxed{0} \\
\text{CAT} = \boxed{\mathbf{raised\text{-}np\text{-}cat}} \\
\text{SEM} = \boxed{5} \\
\text{QUALIA} = \left[\begin{array}{l}
\mathbf{nomqualia} \\
\text{AGENTIVE} = \mathbf{nomagent} \\
\text{TELIC} = \boxed{\mathbf{verb\text{-}sem}} \\
\text{FORM} = \left[\begin{array}{l}
\mathbf{nomform} \\
\text{ABSOLUTE} = \mathbf{real\text{-}form} \\
\text{RELATIVE} = \mathbf{group}
\end{array}\right] \\
\text{CONSTITUENCY} = \left[\begin{array}{l}
\mathbf{groupconst} \\
\text{PARTICLES} = \mathbf{string} \\
\text{ELEMENTS} = \boxed{8} \\
\text{UNITQUANTITY} = \mathbf{string}
\end{array}\right] \\
\text{OBJECT-INDEX} = \mathbf{entity}
\end{array}\right]
\end{array}\right]
\end{array}\right]
$$

Figure 7: The `group-to-plural` type-shifting rule

$$
\left[\begin{array}{l}
\textbf{morph-rule} \\
0 = \left[\begin{array}{l}
\textbf{lex-noun-sign} \\
\text{ORTH} = \left[\begin{array}{l} \textbf{complex-orth} \\ \text{ORTH1} = \boxed{7}\ \textbf{orth} \\ \text{ORTH2} = +\,s \end{array}\right] \\
\text{CAT} = \boxed{\textbf{noun-cat}} \\
\text{SEM} = \left[\begin{array}{l} \textbf{obj-noun-formula} \\ \text{IND} = \boxed{0} \left[\begin{array}{l} \textbf{obj} \\ \text{AGR} = \left[\begin{array}{l} \textbf{agr} \\ \text{NUM} = \textbf{pl} \end{array}\right] \end{array}\right] \\ \text{PRED} = \boxed{1}\ \textbf{logical-pred} \\ \text{ARG1} = \boxed{0} \\ \text{PLMOD} = \textbf{true} \\ \text{QUANT} = \textbf{false} \end{array}\right] \\
\text{QUALIA} = \left[\begin{array}{l} \textbf{nomqualia} \\ \text{AGENTIVE} = \boxed{2}\ \textbf{nomagent} \\ \text{TELIC} = \boxed{3}\ \boxed{\textbf{verb-sem}} \\ \text{FORM} = \left[\begin{array}{l} \textbf{nomform} \\ \text{RELATIVE} = \textbf{plural} \end{array}\right] \\ \text{CONSTITUENCY} = \boxed{5}\ \boxed{\textbf{nomconst}} \\ \text{OBJECT-INDEX} = \boxed{0} \end{array}\right]
\end{array}\right] \\
1 = \left[\begin{array}{l}
\textbf{lex-noun-sign} \\
\text{ORTH} = \boxed{7} \\
\text{CAT} = \boxed{\textbf{noun-cat}} \\
\text{SEM} = \left[\begin{array}{l} \textbf{obj-noun-formula} \\ \text{IND} = \boxed{6} \left[\begin{array}{l} \textbf{obj} \\ \text{AGR} = \left[\begin{array}{l} \textbf{agr} \\ \text{NUM} = \textbf{number} \end{array}\right] \end{array}\right] \\ \text{PRED} = \boxed{1} \\ \text{ARG1} = \boxed{6} \\ \text{PLMOD} = \textbf{false} \\ \text{QUANT} = \textbf{true} \end{array}\right] \\
\text{QUALIA} = \left[\begin{array}{l} \textbf{nomqualia} \\ \text{AGENTIVE} = \boxed{2} \\ \text{TELIC} = \boxed{3} \\ \text{FORM} = \boxed{\textbf{nomform}} \\ \text{CONSTITUENCY} = \boxed{5} \\ \text{OBJECT-INDEX} = \boxed{6} \end{array}\right]
\end{array}\right]
\end{array}\right]
$$

Figure 8: Rule for plural formation