A Comparative Evaluation of UNL Participant Relations using a Five-Language Parallel Corpus

Brian Murphy and Carl Vogel

Brian.Murphy@cs.tcd.ie*, Vogel@cs.tcd.ie
Department of Computer Science, University of Dublin, Trinity College

Abstract. In this paper we describe a manual case study in interlingual translation among five languages. Taking the UN Declaration of Human Rights in Chinese, English, German, Irish and Spanish, we annotated the five texts with a common interlingual logical form. We then studied four inventories of semantic roles (developed for both theoretical and NLP applications), including a subset of UNL's relations, and evaluated their suitability to describe the predicate-argument relationships found in the annotation. As a result, we make some suggestions for possible additions to the UNL relations, and propose that some of the existing relations be conflated or redefined.

1 Introduction

The work described here is part of a feasibility study on the use of semantic roles in interlingua-based machine translation. Our objective was to see if any set of semantic roles could give a description of verb-predicate relationships across a range of languages that would form an adequate basis for automatic generation.

The languages chosen were those that the authors have some working knowledge of (English, Chinese, German, Irish and Spanish), and include widespread and minority languages, both well and less-studied. The corpus used is the UN Declaration of Human Rights [1], a short text covering a broad range of topics in many languages (see Sect. 2).

From the literature on roles we selected four inventories (of which UNL's relations is one) that we considered to be well-enough developed for the annotation of unrestricted text. These inventories ([2,3,4,5] detailed in Sect. 4) were also chosen to be representative both theoretically and in terms of application to tasks such as machine translation and information retrieval.

After aligning the five language versions of the corpus, we manually annotated each article of the text with a language-neutral logical form (effectively a prototype interlingua) following the guidelines described in Sect. 3.1. The main part of the work then involved applying each of the role inventories in turn to the logical form and determining whether they satisfied three key criteria: coverage, differentiation and lack of ambiguity (Sect. 5). In other words, one should

---

be able to annotate every predicate argument with a role, that role should be unique with respect to its predicate, and the assignment of that role should be unequivocal. During this process we also gathered some impressions on the relative strengths and weaknesses of each of the inventories studied.

We had some problems interpreting UNL's documentation on some relations (based on the publicly available specifications and manual [5,6]), and we make suggestions on where this can be improved. In particular we suggest some redefinition of the causal/affected relations AGT, OBJ and AOJ (Sect. 5.1) and propose a more radical rationalisation of the locational relations PLC, PLF/PLT, SRC/GOL and FRM/TO (Sect. 5.2). Finally we consider whether further dedicated relations should be added for arguments that do not contribute as much to causality or directionality, such as possessors/possessions and the peripheral participants of beneficiary and recipient (Sect. 5.3).

2   The Corpus

For our research we were interested in a source of parallel texts that, besides three major Western European languages (German, English and Spanish), included both a minority language (Irish Gaelic) and a major non-European language (Standard Chinese). This rules out most collections from international organisations like the UN or the EU. However the UN Declaration of Human Rights [1], though short (approx. 1500 words) is freely available from the web, and professionally translated to more than 300 languages. While the register is restricted, it covers a wide range of topics, including education, politics, religion, law, the family, asylum, ownership, employment, leisure, culture and health. It offers complex sentence structures (such as deeply nested clauses) and widespread inter-sentence relationships, such as anaphora and mutual conditions between propositions (for example the dependencies between predicates in (2f)), but is simple to align to a sentence level, due to its organisation into articles and sub-articles.

The five languages included cover several branches of the Indo-European family of languages (Celtic, Romance and Germanic) together with a Sino-Tibetan language, and are varied in terms of argument structure. Compared to the fixed *subject-verb-object* structure found in English, German differs in using case and in allowing object fronting, while Spanish allows both object fronting and subject omission (pro-drop). Irish has a *verb–subject–object* word order, while Chinese argument realisation is very flexible, in principle allowing any argument to be moved or dropped. German clause structure differs from the others in grouping non-finite verbs at the end of a clause (e.g. 'gemacht werden' in (1a)), while subordinate clause ordering in Chinese is radically different with modifiers generally preceding heads ('type' and 'right' in (1b)).[1] The copula ('be') has multiple realisations in both Spanish ('ser'/'estar') and Irish ('bí'/'is'), and Irish also has widespread use of prepositional and adverbial forms for representing events – e.g. in (1c) an abstract possession is expressed as being 'at' the owner.[2]

---

[1] DE is a modifier particle.

[2] Examples from the Declaration indicate the source article.

(1)   a.   Berufsschulunterricht müssen allgemein verfügbar gemacht werden ...
              vocation-lesson must general available made to-be ...
              'Professional education shall be made generally available ...' [Art. 26.1]

       b.   父母对其子女所应受的教育的种类,有优先选择的权利
              fùmǔ duì qí zǐnǔ suǒ yīng shòu de jiàoyù de zhǒnglèi, yǒu yōuxiān xuǎnzé de quánlì
              parent to its children that should receive DE education DE type, has priority select DE right
              'Parents have a prior right to choose the kind of education that shall be given to their children.' [Art. 26.3]

       c.   Is ionann na cearta atá acu ...
              is same the rights that-are at-them ...
              'They are entitled to equal rights' [Art. 16]

3   Interlingual Annotation

We manually aligned all 49 articles and sub-articles of the UN Declaration across the five languages, before adding English glosses (i.e. word-for-word translations, as seen in the previous examples) for all the non-English texts. The logical annotation of each article then proceeded on the basis of the English original and the four glosses, yielding over 500 predicates with almost 900 arguments. The aim was to arrive at a single, cross-linguistic logical form that, to the extent possible, adequately represented an article's meaning as expressed in all five versions. Although the result does not follow any of the five surface forms exactly, we aimed to abstract away from them only to the extent necessary to find a common representation.

To our knowledge there are no generally accepted guidelines for the manual annotation of unrestricted text with logical forms, as they are often theory or application specific. However, two sources proved useful. The Penn Propbank (a semantically annotated corpus) guidelines [7] have useful suggestions that we adopted for the treatment of phrasal verbs, support verbs and nominalizations. From cognitive science, Kintsch [8] gives an brief overview of annotation conventions for the 'microstructure' (roughly intra-sentence structure) of propositions, as used in comprehension modelling. We have broadly followed his treatment of negatives, modals, adjectives, adverbs and the status of propositions as arguments themselves.

3.1   Guidelines Developed

Negatives, modal verbs, adjectives and adverbs are expressed as one-place predicates with an event or object argument. As the focus of our studies is valency patterns, the quantification of objects was not annotated and noun phrases are rarely decomposed. Thus "all the rights and freedoms [Art. 2]" would be rendered as the atomic object *AllTheRightsAndFreedoms* as opposed to a form like $[\forall x.[\text{right}(x) \lor \text{freedom}(x)]]$. Tense and aspect are not encoded.

Passive sentences are expressed actively with an undefined logical subject (annotated *U*). Complex sentences are decomposed into component predicates, and nominalizations are given predicate translations where possible (e.g. "interference in privacy" becomes *interfere(U,Privacy)*). Repeated objects and events are given numbered *O* and *E* variables to indicate identity:[3]

(2) a. Everyone charged with a penal offence has the right to be presumed innocent until proved guilty ... [Art. 11.1]

  b. 凡受刑事控告者,在... 证实有罪以前,有权被视为无罪...

    fán shòu xíngshì kònggào zhě, zài ... zhèngshí yǒu zuì yǐqián, yǒu quán bèi shìwéi wúzuì

    every receive criminal charge person, at ... confirm has guilt before, has right BEI regard innocent

  c. Jeder, der wegen einer strafbaren Handlung beschuldigt wird, hat das Recht, als unschuldig zu gelten, solange seine Schuld nicht ... nachgewiesen ist ...

    everyone, who because-of a criminal act charged be, has the right, as innocent to count, while his/her guilt not proved is

  d. Gach duine a cúiseofar i gcion inphíonois is tuigthe é a bheith neamh-chiontach go dtí go gcruthaítear ciontach é ...

    every person that charged in offence punishable be understood him that be innocent until that prove guilty him

  e. Toda persona acusada de delito tiene derecho a que se presuma su inocencia mientras no se pruebe su culpabilidad ...

    every person accused of crime has right to that one presumes his/her innocence while not one proves his/her guilt

  f. E1:charge(U1,O1:Anyone,E2:penally(offend(O1)))
    depend(E3,not(E4)) E3:entitled(O1,presume(U3,innocent(O1)))
    E4:prove(U2,guilty(O1,E2))

Support verb constructions (e.g. 'give education', 'subject to limitations' etc.) are reduced to their nominal object as predicate. Thus the meaning of "enjoy ... protection" is expressed with the predicate *protect()*:

(3) a. All children ... shall enjoy the same social protection. [Art. 25.2]

  b. 一切儿童...都应享受同样的社会保护

    yīqiē értóng ... dōu yìng xiǎngshòu tóngyàngde shèhùi bǎohù

    all child ... all should enjoy same society protect

  c. Alle Kinder ... genießen den gleichen sozialen Schutz

    all children ... enjoy the same social protection

  d. Bhéarfar an chaomhaint shóisialach chéanna don uile leanbh ...

    given the protection social same to all children ...

  e. Todos los niños ... tienen derecho a igual protección social

    all the children ... have right to equal protection social

  f. shall(equally(protect(U1,AllChildren)))

---

[3] In all examples from the corpus languages are listed in the following order: English, Chinese, German, Irish, Spanish. BEI is an agentive marker.

Many of the conflicts between annotations suggested by individual language glosses are superficial, (e.g. near synonyms such as 'fair' (English) versus 'córa' (Irish: 'just') and 'equitativo' (Spanish: 'equitable')), in which case one of the lexicalisations is arbitrarily chosen. However, when there is a conflict in meaning we use two criteria to decide on a common predicate structure. Majority rule is one – for example in (2), the predicate 'presumed' won out, as it is used in both Spanish and English, and we judged it semantically close to 'regarded' (Chinese) and 'understand' (Irish), but significantly different from the German 'count'. Secondly, subject to majority rule, the most componential logical form available is used, as what is lexicalized in one language as a single verb may be a verb-argument complex in another. Hence, in the example below, the form *expel(U,Person,Country)* as suggested by the German version is preferred over *exile(U,Person)*.

(4)   a.   No one shall be subjected to . . . exile [Art. 9]
        b.   任何人不得加以. . . 放逐
             rènhé rén bùdé jiāyǐ . . . fàngzhú
             any person must-not be-made . . . exile
        c.   Niemand darf . . . des Landes verwiesen werden
             no-one may . . . the country expelled be
        d.   Ní déanfar . . . aon duine . . . a chur ar deoraíocht
             not make . . . single person . . . that put in exile
        e.   Nadie podrá ser . . . desterrado
             no-one will-be-able to-be . . . exiled
        f.   shall(not(expel(U,O1:Anyone,O2:Country))) belong(O1,O2)

We have not yet settled on semantic model of the formal language we use, but it resembles a higher-order logic, or a first-order logic with named Skolem functions.

4   Models of Semantic Roles

Semantic roles were first posited by linguists to describe the nature of meaning relationships among arguments and verbs in sentences. They correspond to a subset of UNL's relations. In this work we concentrate on so-called participant relations (see Table 1) as opposed to the more oblique circumstantial roles such as *manner, purpose* or *condition*, which are less commonly included in role inventories.

The earliest role inventories [9,10] were causally based and mirrored the grammar of argument structure quite closely (consider Fillmore's *agentive, dative,* and *objective* cases). Jackendoff went on to introduce a localist hypothesis [11] (or "thematic hypothesis") based on the extension of verbs (e.g. 'stay', 'go') and prepositions (e.g. 'from', 'to', 'at') of location and movement to more abstract situations (5). For example, information is viewed as *theme* ('story' in (5d) and by extension 'what' in (5g)) and holders can be viewed as *location* ('student' in (5d) and by extension 'document' in (5e) and 'mine' in (5f)).

Table 1. Typical participant roles

Agent	the (typically animate) volitional initiator of an action
Effector	the non-volitional initiator of an action
Patient	the affected party, or undergoer of the action
Theme	the entity whose state, movement or location is described
Experiencer	the entity that perceives the situation
Percept	the entity that is perceived
Recipient	the entity to which another entity is passed
Beneficiary	the entity to whose advantage the action is performed
Instrument	the entity with which the action is performed
Goal	the location towards which an entity moves
Source	the location away from which an entity moves

(5) a. Ciara$_{theme}$ stayed [at work]$_{location}$/[angry]$_{location}$
 b. Saoirse$_{theme}$ went [from A$_{source}$ to B$_{goal}$]/[from happy$_{source}$ to sad$_{goal}$]
 c. The meeting$_{theme}$ will be at [the main office]$_{location}$/[6pm]$_{location}$
 d. The teacher$_{source}$ [gave]/[told] a story$_{theme}$ to her student$_{goal}$
 e. Your ideas$_{theme}$ were not included in the document$_{location}$
 f. The tricycle$_{theme}$ is mine$_{location}$!
 g. They$_{location}$ know what$_{theme}$ they're talking about

There are obvious problems with both the purely causal or localist approaches. It is unclear how a localist scheme would tag an *instrumental* role, and with verbs of perception (e.g. 'hear', 'look') is the *experiencer* the *goal* or the *source*? Nor is it obvious how a purely causal scheme would distinguish between spatial *source* and *goal* (e.g. (5b)).

4.1 Hybrid Models of Roles

Because of these difficulties Jackendoff developed a hybrid, two-tier scheme [2] as part of his semantic representation (Lexical Conceptual Structure, orLCS) with his localist roles on a 'thematic' tier, and causal roles in an orthogonal 'action' tier:

(6)   Pete$_{source\&agent}$ kicked the ball$_{theme\&patient}$ down the field$_{goal}$

Saeed [12] suggests completing the Jackendoffian scheme as such, and it is this version that we use here (*actor* is equivalent to *effector*):

Thematic Tier theme, goal, source, location
Action Tier actor, agent, experiencer, patient, beneficiary, instrument

Dorr [3] took Jackendoff's work as a departure point when designing a semantic representation for the lexicon of her interlingual machine translation system UNITRAN, also seeking to "...strike a balance between the causal and motion/location dimensions ...". Her inventory differs in being on a single tier and

by incorporating situation-specific roles such as *information* and *percept* (Table 2). She has made an extensive verb lexicon available [13], where each of the 11 thousand entries is annotated with argument syntax and role structure, using verb frames based on Levin's [14] semantic classes.

Table 2. Dorr's LCS roles

AG	agent	**TH**	theme
EXP	experiencer	**INFO**	information
SRC	source	**GOAL**	goal
PERC	perceived item	**PRED**	identificational predicate
LOC	locational predicate	**POSS**	possessional predicate
BEN	benefactive modifier	**ISNTR**	instrument modifier
PROP	event or state	**PURP**	purpose modifier or reason
MANNER	manner	**TIME**	time modifier

Sowa [4] has developed a model of roles for knowledge representation (see Table 3) based on Dick's [15] work in information retrieval, and Somers' Case Grid [16]. Sowa replaces the locational column labels (*Source, Path, Goal, Local*) of Somers and Dick with the four causes from Aristotle's *Metaphysics* (*Initiator, Resource, Goal, Essence*) and introduces six intuitive verb classes, which combined with several additional distinguishing features (such as animacy for differentiating *agent* and *effector*) correspond to more conventional roles.

Table 3. Sowa Roles

	Initiator	Resource	Goal	Essence
Action	Agent, Effector	Instrument	Result, Recipient	Patient, Theme
Process	Agent, Origin	Matter	Result, Recipient	Patient, Theme
Transfer	Agent, Origin	Instrument, Medium	Experiencer, Recipient	Theme
Spatial	Origin	Path	Destination	Location
Temporal	Start	Duration	Completion	PointInTime
Ambient	Origin	Instrument, Matter	Result	Theme

The model of relations used by UNL is more extensive, including logical operators such as AND and OR, and other novel roles such as BAS (*basis for expressing degree*) and SEQ (*sequence*). In particular it gives us a comprehensive treatment of the commitative roles CAG, COB and CAO (*co-agent, affected co-thing* and *co-*

thing with attribute) not offered by any of the other schemes examined. They allow us to express the difference in focus between (7a) and (7b):

(7)   a. [Fergal and Fergus]$_{OBJ}$ bumped into each other on the street
        b. Fergus$_{OBJ}$ bumped into Fergal$_{COB}$ on the street

5   Comparative Evaluation of UNL Relations

The following evaluation is essentially critical, in that we draw attention only to shortcomings of UNL relations or their documentation. The treatment given here of the other inventories will not be comprehensive – rather we will mention them only where they seem to provide a superior solution to UNL. To make a fair comparison, the assignment of roles was carried out on the interlingual form described in Sect. 3, rather than in the context of the semantic representation intended for each inventory (i.e. UNL Expressions, Sowa's Conceptual Graphs, or LCS for the Jackendoff and Dorr schemes). The criteria we used for evaluating role assignments were as follows:

1. Coverage: must be able to assign a role to every argument of every predicate, e.g. in (5b) we saw how a purely causal scheme would fail to express spatial start and end points
2. Differentiation: must be able to assign a unique role to every argument with respect to its predicate, e.g. a scheme without commitative roles would lack differentiation between the syntactic subject and object in (7b)
3. Lack of Ambiguity: must be able to assign a single role unequivocally to each argument - an argument should not fit multiple roles or fall between roles, e.g. a single tier scheme might be unclear on whether 'ball' in (6) is a *theme* or *patient*

Generally, all four inventories performed well on coverage and differentiation, though Jackendoff's small number of roles sometimes presented problems of duplicate assignments to single predicates. Most problems we encountered were with ambiguity. In the following discussion, we make suggestions for alterations to the UNL relations according to the principle that they should adequately and efficiently express generalisations either in semantics (e.g. inferences that can be drawn) or in syntax (e.g. structures that are licensed). We now examine some problematic aspects of the UNL relations in turn, based on the definitions and prototypical examples given in [5,6].

5.1   Causal Relations: AGT, OBJ, AOJ, INS

AGT agent: thing that initiates an action, e.g. "John$_{AGT}$ broke the window"
OBJ affected thing: thing in focus which is directly affected by an event or state, e.g. "write a novel$_{OBJ}$"
AOJ thing with attribute: thing which is in a state or has an attribute, e.g. "This flower$_{AOJ}$ is beautiful"

INS instrument: instrument to carry out an event, e.g. "cut with scissors$_{INS}$"

The OBJ relation (i.e. *patient* role) is used for both clearly affected patients (e.g. the *Anyone* argument of *expel()* in (4f)) and for less affected participants such as the complements of psychological verbs (e.g. the *innocent()* argument of *presume()* in (2f)) and communication verbs (e.g. 'story' in (5d)). While this in itself may not be a problem, it may be missing significant syntactic generalisations. In several languages the tendency of a syntactic object to be promoted to a more prominent position, such as subject, seems in part determined by its affectedness. In the examples below the passive (8b) and 'ba'/'bei' (9b, c) variants of *enjoy(I,TheArts)*, all of which promote the object, are anomalous:[4]

(8)   a.   I$_{AGT}$ enjoy the arts$_{OBJ}$ [variation on Art. 27.1]
       b.   * the arts$_{OBJ}$ get enjoyed by me$_{AGT}$
(9)   a.   我享受艺术
             wǒ$_{AGT}$ xiǎngshòu yìshù$_{OBJ}$ [Chinese]
             me enjoy art
       b.   * 艺术被我享受
             * yìshù$_{OBJ}$ bèi wǒ$_{AGT}$ xiǎngshòu
             art BEI me enjoy
       c.   * 我把艺术享受
             * wǒ$_{AGT}$ bǎ yìshù$_{OBJ}$ xiǎngshòu
             me BA art enjoy

Both [2] and [4] give a directional interpretation of these verbs, where the *enjoyer* above is a *goal* and 'the arts' a *source*. However examples from our corpus show that using the localist hypothesis (see Sect. 4) with these verbs does not generalise across languages. As we see below (10), in German our enjoyment is 'in' the arts, while in Irish almost the reverse is true – the enjoyment is 'at' us. As a result we suggest that a simple alternative is to use AOJ (roughly equivalent to *theme*) for non-affected syntactic objects. A more significant reworking would be to add the new roles of PRC (*percept*) and INF (*information*) following the practise of [3].

(10)   a.   Everyone$_{AGT}$ ... to enjoy the arts$_{OBJ}$ ... [Art. 27.1]
        b.   ... sich$_{AGT}$ an den Künsten$_{OBJ}$ zu erfreuen ... [German]
              ... self at the arts to enjoy ...
        c.   ... áineas na n-ealaíon$_{OBJ}$ a bheith aige$_{AGT}$ ... [Irish]
              ... pleasure of-the arts that be at-him ...
        d.   enjoy(Everyone,TheArts)

---

[4] A 'got' passive is used here as it cannot be mistaken for a non-passive adverbial sentence such as "he was unimpressed by the play". The star '*' indicates an idiosyncratic or ungrammatical form. The relation annotations shown follow UNL as it stands, rather than our proposals. BEI is an agentive marker and BA is an affectedness marker, both of which promote the object to a preverbal position.

Similarly there is a tendency for non-volitional or inanimate subjects (such as "I$_{AGT}$ think", "someone$_{AGT}$ is sleeping" and "a process$_{AGT}$ makes something") to resist being demoted by passivisation or other processes. We suggest that INS could be used for inanimate initiators such as 'a process', or a new EFT (*effector*) relation could be introduced (see [2,4]). The subjects of psychological verbs (e.g. 'think' and 'sleep' above) could take the AOJ relation, or a newly coined EXP (*experiencer*) relation. However, then we would lose the distinction between the volitional and non-volitional subjects of perception verbs such as 'listen'/'hear' and 'watch'/'see' – the relative merits are debatable.

5.2   Locational Relations: PLF/PLT, SRC/GOL, FRM/TO, PLC

PLF   initial place: the place an event begins or a state becomes true, e.g. "come from home$_{PLF}$"

PLT   final place: the place an event ends or a state becomes false, e.g. "leave for India$_{PLT}$"

SRC   initial state: initial state of object or the thing initially associated with object of an event, e.g. "the light changed from red$_{SRC}$"

GOL   final state: final state of an object or the thing finally associated with an object of an event, e.g. "getting better$_{GOL}$"

FRM   origin: origin of a thing, e.g. "a letter from him$_{FRM}$"

TO   destination: destination of a thing, e.g. "a train to Edinburgh$_{TO}$"

PLC   place: place an event occurs or a state is true or a thing exists, e.g. "stay at home$_{PLC}$"

FRM/TO are problematic as they are used for two rather different purposes: describing the concrete path a *Thing* takes, as in the *Country* argument of the *expel()* predicate in (4f); and for the origin of a *Thing*, as seen in *belong()* of the same example. These two functions are treated quite differently in three of the languages examined. Consider possible translations for the constructed examples "the man from London" and "the train from London" respectively:

(11)   a.   伦敦的人/ 伦敦来的火车 [Chinese]
             lúndūn de rén / lúndūn lái de huǒchē
             london DE person / london come DE train

       b.   an fear as London / an traen ó London [Irish]
             the man out-of london / the train from london

       c.   el hombre de Londres / el tren desde Londres [Spanish]
             the man of london / the train from london

While "lúndūn de huǒchē" and "el tren de Londres" are both possible, they can mean several things, including the train both going to or coming from London, much as "the London train" can in English. As predicates exist for some other prepositions, for example *against()*, we suggest using a new predicate called *origin(*AOJ,PLC*)* for describing the provenance of a thing.

For concrete path uses of FRM and TO, we suggest that these relations be conflated with PLF/PLT. None of the other role inventories examined have an

event/entity distinction when it comes to locational roles, and the UNL relation PLC can be applied to both *Things* and *Events* (e.g. "a town$_{Thing}$ in Bavaria$_{PLC}$" and "She is$_{Event}$ in Bavaria$_{PLC}$"). In addition, it seems strange that the English prepositions 'from' and 'to' receive such special treatment, while the similarly common 'in' and 'of' do not.

Initially, the opposition of PLF/PLT for locations (e.g. (12) "return to his country$_{PLT}$") with SRC/GOL for states (e.g. (1a) "make education available$_{GOL}$") seems well justified.

(12) a. Everyone has the right ... to return to his country [Art. 13.2]
      b. 人人有权...返回他的国家
          rénrén yǒu quán ... fǎnhuí tā de guójiā
          everyone has right return s/he DE country
      c. Jeder hat das Recht ... in sein Land zurückzukehren
          everyone has the right in his/her land to-return
      d. Tá ag gach uile dhuine an ceart chun ... filleadh ar a thír féin
          is at each every person the right to return to his country own
      e. Toda persona tiene derecho ... a regresar a su país
          every person has right to return to his/her country
      f. entitled(O1:Everyone,return(O1,O2:Country)) belong(O1,O2)

However, some of the examples given in the documentation blur the distinction, in particular "go to Brussels$_{GOL}$" and "withdraw from the stove$_{SRC}$". It is not clear to us what basis there is for differentiating between 'his country' above as the final state of the entity 'Everyone' (GOL) or the final place of the event 'return' (PLT) – in both cases the ending of the event and the arrival of the agent happens in the same place at the same time. As a result, we suggest restricting SRC/GOL to non-spatial states only.

A more radical alternative would be to eliminate the SRC/PLF and GOL/PLT distinction altogether. We do not make a similar distinction for static locations (stative "famous in his field" and spatial "live here" both use PLC), and this is supported by [2,3] where spatial and stative end-points are conflated in *source/goal*.

5.3   Miscellaneous: POS, BEN

POS possessor: possessor of a thing, e.g. "the company's$_{POS}$ building"
BEN beneficiary: not directly related beneficiary or victim of an event or state,
    e.g. "be fortunate for you$_{BEN}$"

Possession is treated differently in UNL, depending on whether a genitive form ("that is my car$_{POS}$") or a possessional predicate ("I$_{AGT}$ have a pen$_{OBJ}$") is used. As with FRM/PLF and TO/PLT this seems like an unnecessary complication that none of the other inventories require. We also have to ask how agentive the subjects of verbs like 'have' and 'own' are – e.g. in what sense is the subject of "I have no money" an *agent*? Again we see that sentences of this type resist passivisation in English (13b) and the 'ba'/'bei' constructions in Chinese (14b, c). We suggest that possession be annotated as *possess(*POS,AOJ*)* following the practise of [3].

(13) a. . . . [people$_{AGT}$] own property$_{OBJ}$ . . . [variation on Art. 17.1]
     b. * property$_{OBJ}$ gets owned by people$_{AGT}$
     c. own(People,Property)

(14) a. 人所有财产
         rén$_{AGT}$ suǒyǒu cáichǎn$_{OBJ}$
         people own property
     b. * 财产被人所有
         * cáichǎn$_{OBJ}$ bèi rén$_{AGT}$ suǒyǒu
         property BEI people own
     c. * 人把财产所有
         * rén$_{AGT}$ bǎ cáichǎn$_{OBJ}$ suǒyǒu
         people BA property own

The beneficiary relation BEN works well for adjuncts in English (e.g. "do something for you$_{BEN}$"), but we suggest it be extended to beneficiary syntactic objects. These are currently assigned the GOL relation, even though "make someone$_{GOL}$ a cup of tea" is equivalent to "make a cup of tea for someone$_{BEN}$".

In our opinion a *recipient* relation REC is also needed [3,4]. Note how in English *recipient* arguments ('Anja' in constructed example (15a)) can be syntactic objects, while inanimate arguments that would take a GOL or PLF relation (e.g. 'Munich') cannot – rather an adjunct is necessary, as in "I sent a present to Munich". In German different prepositions and case are used to express these two roles (accusative 'an' for *recipients* and dative 'nach' for *goals*).

(15) a. I sent Anja/*Munich a present
     b. Ich habe ein Geschenk an Anja/nach München geschickt
         I have a present to Anja/to Munich sent
     c. send(I,Present,Anja) / send(I,Present,Munich)

6   Conclusion

In this work a prototype interlingua was manually applied to a five-language parallel corpus to reveal predicate valency patterns. Then several inventories of semantic roles, including a subset of UNL's relations, were assigned to the resulting logical forms. In the subsequent evaluation UNL performed well in terms of coverage and differentiation, but we encountered some problems of ambiguity in the assignment of locational and causal relations. As a result we have some opinions on how parts of the UNL relations might be reformed, based on semantic and syntactic generalisations in the languages examined (English, Chinese, German, Irish and Spanish) and particular structures we encountered in the corpus.

Firstly, we propose that the FRM/TO relations be folded into PLF/PLT, and that the distinction between spatial end-points PLF/PLC and stative end-points SRC/GOL be firmed up. We also propose redrawing the lines between causal relations – specifically non-affected *objects* (e.g. the syntactic objects of communication verbs) should be assigned AOJ rather than OBJ, and non-volitional *agents* should be assigned INS rather than AGT. We propose extending the usage

of BEN from adjuncts to also cover syntactic objects, and using POS for verbal as well as nominal structures that express possession. Finally we suggest several new situation specific roles (*recipient, effector, experiencer*) and explain how they might be of use in future versions of UNL.

References

1. United Nations General Assembly: Universal declaration of human rights. http://www.unhchr.ch/udhr/navigate/alpha.htm (1948) [Viewed December 2004].
2. Jackendoff, R.: Semantic Structures. MIT Press, Cambridge (1990)
3. Dorr, B.J.: Machine Translation: A View from the Lexicon. MIT Press, Cambridge (1993)
4. Sowa, J.F.: Knowledge Representation: Logical, Philosophical, and Computational Foundations. Brooks/Cole, London (2000)
5. UNL Centre: The universal networking language specifications 3.2. http://www.undl.org/unlsys/unl/UNL Specifications.htm (2003) [Viewed December 2004].
6. UNL Centre: UNL manual. http://www.undl.org/unlsys/unlman/index.html (2001) [Viewed December 2004].
7. Kingsbury, P.: Propbank annotation guidelines. http://www.cis.upenn.edu/∼ ace/propbank-guidelines-feb02.pdf (2002) [Viewed December 2004].
8. Kintsch, W.: Comprehension: A Paradigm for Cognition. Cambridge University Press, Cambridge (1998)
9. Gruber, J.S.: Studies in Lexical Relations. Indiana University Linguistics Club, Bloomington (1965) Reprint of PhD Thesis.
10. Fillmore, C.J.: The case for case. In Bach, E., Harms, R.T., eds.: Universals in Linguistic Theory. Holt, Rinehart and Winston, New York (1968) 1–92
11. Jackendoff, R.: Semantic Interpretation in Generative Grammar. MIT Press, Cambridge (1972)
12. Saeed, J.I.: Semantics. Blackwell, Oxford (1997)
13. Dorr, B.J.: LCS database documentation. http://www.umiacs.umd.edu/∼ bonnie/LCS_Database_Documentation.html (2001) [Viewed December 2004].
14. Levin, B.: English Verb Classes and Alternations. University of Chicago Press, Chicago (1993)
15. Dick, J.: A Conceptual, Case-Relation Representation of Text for Intelligent Retrieval. PhD thesis, Department of Computer Science, University of Toronto (1991)
16. Somers, H.: Valency and Case in Computational Linguistics. Edinburgh University Press, Edinburgh (1987)