

AMI Named Entity Guidelines

Version 1.0

Some person
person@dfki.de

Vasilis Karaiskos
vkaraisk@inf.ed.ac.uk

02/05/2005

The scheme follows the NIST “1999 Named Entity Recognition Task Definition”, version 1.4, (HUB4-NE, http://www.nist.gov/speech/tests/ie-er/er_99/doc/ne99_taskdef.v1.4.ps). Section 3 of that document does not apply to this task.

The points where our scheme deviates or extends NIST 1999 are detailed below. The hierarchical structure of the scheme is shown in fig. 1. Note that only tree leaves can actually be used as entity labels.

1 PERSON

The PERSON tag has been subdivided into categories that reflect the particular needs of the corpus.

1.1 PARTICIPANT

This category comprises four tags corresponding to the roles allocated to the meeting participants by the scenario. Naturally, there may be slight variations in the actual wording used by speakers. An entity should be marked as long as it thought to be one, regardless of such differences in form.

1. PROJECT_MANAGER (p_pm)

I am (p_pm. Betty). I am the (p_pm. project manager) for today.

2. INTERFACE_SPECIALIST (p_is)

Uh sure my name is (p_isAgnes) and I'm an user usability (p_is.user interface designer).

3. MARKETING (p_market)

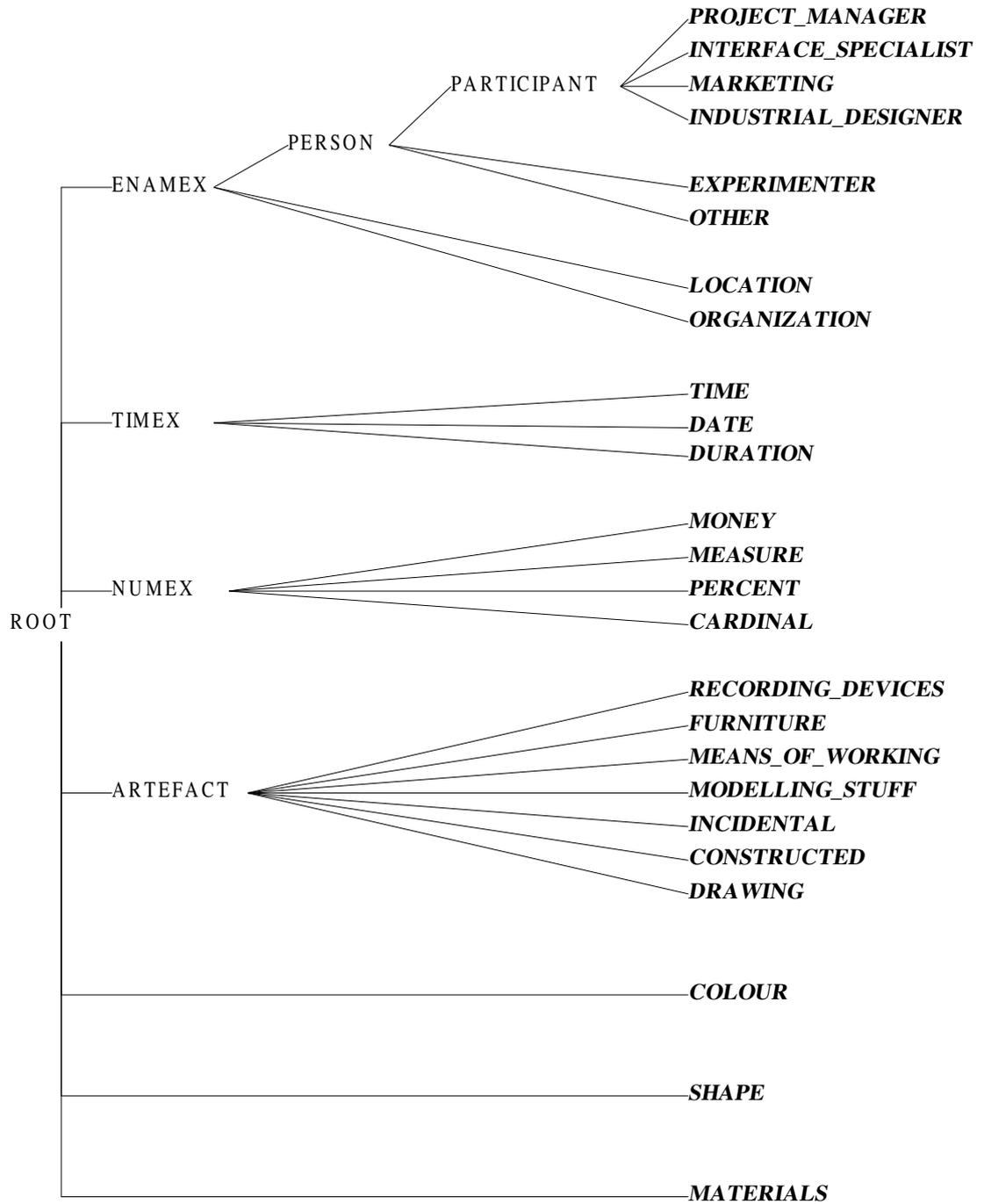


Figure 1: Named entity hierarchy.

I'm a (p_market. marketing person)

4. INDUSTRIAL_DESIGNER (p_id)

Uh I'm (p_id. Jean-Oui). Um uh my role is (p_id. industrial designer)

1.2 EXPERIMENTER (p_exp)

This tag is used to mark references to anyone on the research team who gives instructions to the participants or sets up the recording equipment.

instructions will be sent to you by your p_exp(coaches)

This should include any person who enters the meeting room; it is generally not expected that other people will be present at any time.

1.3 OTHER_PERSON (p_other)

This tag applies to any other person referred to by the participants.

this guy p_other(Zenityh) created the Flashmatic

1.4 Differences from NIST 1999

In NIST 1999 only proper names can be annotated. Here, however, the scheme is extended to include other name-type designators for individuals. In the scenario case this usually means all mentions of the scenario roles.

*I'm p_is(Craig) and I'm p_is(User Interface)
Since I'm not really (p_id. Industrial Designer), I didn't really
know what to do*

Similarly, common nouns can be marked as OTHER_PERSON only in those cases where they uniquely identify a particular person:

It was suggested to me by the (p_other. boss) at my work.

1.5 Non-taggable PERSONs

Pronouns should never be tagged. Additionally, no improvised roles,

*you are my co-ordinator, am I right?
My compliments to the artist.*

generic expressions,

*A participant can draw their favourite animal
if we get a really good designer
make it more attractive to users*

or collective expressions,

*...a team composed of these people
I need to go to the management and tell them.*

are to be tagged.

2 ARTEFACT

This type does not exist in NIST 1999 at all. In our present scheme it is used to annotate only objects that are (thought to be) physically present. Such entities should be annotated inclusively, i.e. if the coder has doubts whether a candidate entity is “present” or not, they should select it.

*I guess the (mow. tool) is really our the (mow. computer), as
far as I can see*

Terms that express general categories of ARTEFACTs can still be annotated as long as the referent is a physically present object.

I can't write with this (mow. thing).

References to parts of artefacts should still be coded as ARTEFACT; for example, descriptions of either objects or parts of objects in a photograph are equally annotated.

*you can actually see on it are the on off (drw. switch), volume
and channel (drw. control) , the menu access (drw. button)*

This type of markable comprises the following sub-categories:

1. **FURNITURE (furn)**, *e.g. table, curtain, chair, partition*
2. **MEANS_OF_WORKING (mow)** *e.g. whiteboard, pen, notepad, projector, laptop, slides*
3. **RECORDING_DEVICES (rec)** *e.g. microphone, camera, recording tower*
4. **MODELLING_STUFF (mod)** *e.g. plasticine, play-dough; it is used specifically for quantities of materials that are physically present (see sec. 5 for other mentions of materials).*

5. **CONSTRUCTED (const)** *e.g. remote control, button, fascia*; any prototypes or parts of prototypes that the participants have put together.
6. **DRAWING (drw)** *e.g. cat, remote control*; any object that is in a photograph, slide, drawn on the whiteboard etc.
7. **INCIDENTAL (inc)** *e.g. fruit, sandwich, mobile phone*; any other object being in the room including bought remote controls that the participants may have with them.

Note that an ARTEFACT can be annotated in different ways depending on the particular context: a remote control can potentially be INCIDENTAL, CONSTRUCTED or a DRAWING.

2.1 Scope of expression

Only the head of the noun phrase forms the entity. Adjectives and other modifiers should be left out, unless they are part of the head (e.g. the whole expression can be found as a single dictionary entry).

(inc. mobile phone)
(inc. white board)
LCD (const. display)
flip (drw. panel)

2.2 Non-taggable artefacts

As with PERSON, no pronouns should ever be considered markable entities. Additionally, any expressions that are generic or have no referent to a physically present object are not to be tagged.

doing research and taking remote controls and looking
Theoretically with the T.V. you already have a big display right
in front of you.

3 COLOUR (col)

Taggable expressions include basic (*green, yellow*) and elaborate colour terms (*yellowy orange*). Colours do not have to be monolexemic, but they need to be a term as opposed to a description. In general, if the expression includes a colour-denoting adjective, it should be tagged as a whole.

(col. mahogany red)
(col. dark blue)

For completion, when a colour is explicitly defined only by a general property, this property should also be marked.

We should paint the buttons a (col. strong) colour.

A colour should be marked whether it is mentioned as an abstract property or that of a physically present object.

3.1 Scope of expression

If a colour is defined with a simile, prepositional phrase, relative clause, etc. only the the colour term as described above should be tagged.

(col. white) as snow
the (col. blue) of the Mediterranean sky
that kind of (col. purple) that fills the sky at sunset

The word “colour” is not included in the expression.

the colour would be customisable

4 SHAPE (shp)

The criteria to follow are similar to those for COLOUR, although the descriptions tend to be much simpler. A markable can be an identifiable geometric shape

(shp. squarey),
(shp. kidney) shape,
(shp. curved) cases

or an idiomatic use of the notion, in which case the fact that this is a reference to shape has to be explicit:

(shp. friendly) shape, (shp. ergonomic) shape

Similarly to COLOUR, any reference to shape should be marked whether it is an abstract mention or a property of a physically present object. The word “shape” is not included in the expression.

5 MATERIALS (mat)

This category is used for entities referring to substances things are made of. It is only used fro generic mentions of materials (cf. MODELING_STUFF for quantities that are physically present).

our remote control could be made of (mat. wood) .
(mat. Titanium) would be heavy.