

## Annotation Guidelines

**Based off of the following and manipulated for Irish:**

- Universal NER Guidelines (Mayhew)
- Linguistic Data Consortium Annotation Guidelines (Linguistic Data Consortium)

**Tags Discussed:**

Person, Location, Organisation, Event

### 1. Person:

- **A person's name or family name (real or fictional even if spelled incorrectly)** eg Micheál Martin(PER)
- **Gods (when having a single reference and capitalised)** eg Buíochas le Dia(PER)
- **A person's initials** eg M.M.(PER)
- **Do not tag titles as a name or part of a name** eg Chonaic mé Dr. O'Sullivan(PER) inné, An Taoiseach Leo Varadkar(PER), An tUasal Mac Gabhann(PER), Mary Lou McDonald(PER) T.D. **BUT do tag if used as a name mention** eg Dúirt An Taoiseach(PER) go bhfuil...

### 2. Location:

- **Geographical places, facilities or buildings**
- Eg countries, cities, towns, airports, hotels, roads etc.
- **When two locations are consecutive tag separately** eg tá mé i mo chónaí i Sord(LOC), Baile Átha Cliath(LOC)
- **Tag whole postal addresses as one** eg 27 Bóthar na Foraoise, Caisleán an Chomair, Cill Chainnigh(LOC)

### 3. Organisation:

- **Named collections of people (organisations, institutions, firms, political parties, unions, groups)** eg Is é Leo Varadkar(PER) ceannaire Fhine Gael(ORG), Fáiltím roimh an nuacht is deireanaí a chuala muid ar maidin ó Citylink(ORG) go mbeidh níos mó busanna ar fáil ón gClochán(LOC) go Gaillimh(LOC), Tá Seanad Éireann(ORG) tar éis an Bille na hOidhreacht Stairiúla agus Seandálaíochta agus Forálacha Ilghnéitheacha, 2023 a rith gan leasú, Anuas air sin, tá an Rialtas(ORG) seo tar éis maoiniú a chur ar fáil do TG4(ORG) chun an stáisiún nua do pháistí, Cúla4(ORG), a bhunú.
- **Names of places when they act as administrative units or sports teams** eg Chaill Baile Átha Cliath(ORG) in aghaidh Gaillimh(ORG) an seachtain seo caite.
- **Include corporate designators like Co. and Ltd. as part of the name** eg Is gnólacht dlí iad Johnson & Co(ORG)
- **Only tag brands if referring to the organisation itself, not as a brand label** eg Tá bróga nua eisithe ag Nike(ORG) VS Ghortaigh mé mo chosa mar chaith mé Nike.

### 4. Event (not currently tagging for event):

- **Major Public Holidays / Seasons** eg Nollag, Lá Fhéile Pádraig, an Cháisc, an gheimhridh, Oíche Shamhna

- **Natural Disasters when given full title** e.g. Hurricane Katrina, Mount Vesuvius eruption. Do not include e.g. na tuilte i gContae na Gaillimhe (here Galway should be labelled as a location and floods should be ignored)
- **Major Historical events/battles-** Labelled event when they refer to a *specific* moment in history - Éirí Amach na Cásca, an Dara Cogadh Domhanda, Brú na Bóinne. These will be well-known historically significant occurrences in history (will normally be in capitals)
- **Major Sporting Events-** Well known sporting events like - na Cluichí Oilimpeacha, Cluiche Ceannais na hÉireann, an Corn Domhanda (don't include common events like cluiche peile)
- **Do NOT include personal events such as birthdays, weddings etc** - e.g. lá breithe Liam (only Liam should be labelled as PER)
- **Include Major concerts/festivals** - e.g. D'fhreastal mé ar an bhféile - EP .... Do **Not** Include things like “ceolchoirm Taylor Swift”. Taylor Swift should be tagged as PER in this instance

**5. Other points to note:**

- **Inclusion of non-name tokens** eg Nua-Eabhrac-bhunaithe(LOC)

**References:**

Linguistic Data Consortium. *ACE (Automatic Content Extraction) English Annotation Guidelines for Entities*. 13 June 2008.

Mayhew, Stephen. “Annotation Guidelines.” *Universal NER*, Nov. 2022, [www.universalner.org/guidelines/](http://www.universalner.org/guidelines/). Accessed 24 Nov. 2023.