Guidelines for the Annotation of Nominal Utterances in POP-HS-IT

(v. 1.0 July 2019)

GLORIA COMANDINI

Università degli Studi di Trento, Italy

gloria.comandini@unitn.it

This document presents a quick annotation framework of nominal utterances (NUs) on Twitter.

This annotation framework is a simplified version of the original Italian guidelines for annotating NUs, *Linee guida per l'annotazione degli enunciate nominali[1]*, used for the annotation of COSMIANU (Corpus Of Social Media Italian Annotated with Nominal Utterances).

The annotation of NUs should be done on an Excel file; the NUs found on a tweet should be annotated on the same line, but under the column "Tag_NU". The NUs should be preceded by **<NU>** and followed by **</NU>**, separated from the text by a space, like in Table 1.

Table 1

Text	Tag_NU
Non vogliamo più stranieri! Tutti a casa!	<NU> Tutti a casa! </NU>
We don't want more strangers! Everybody home!	<NU> Everybody home! </NU>

NB: this annotation framework is based on the Italian language. If applied to other languages, you should be careful of linguistic differences.

1. Nominal Utterances (NUs)

An utterance is a segment of really produced text of any extension and of any syntactic form and it could consist in one or more clauses. "It rains" is an utterance just like "Someone could believe that the majority of humanity, if examined alone and not in a group, is intrinsically good, as Locke said, since he was notoriously more optimistic than Hobbes".

Both these utterances are verbal, since they both have a verb in a finite form in their main clause (*rains* and *could believe*).

When an utterance doesn't have a verb in a finite form in its main clause, it is called a nominal utterance, like "Good morning!" or "A good day for a long walk in the park".

The annotation of NUs consists in two steps:

    A. Identify the NUs in the text;

---

[1] The original guidelines are available (in Italian) here: http://tiny.cc/auhvvy.

B. Copy-paste the NUs in the Tag_NU column and annotate the NUs with the tags \<NU\> and \</NU\>. \<NU\> signals the beginning of the NU, while \</NU\> closes the annotation at the end of NU's extension.

2. The Annotation of simple NUs and NUs with subordinate clauses

NUs are defined by what they don't have: a verb in a finite form. Thus, a NU can contain every other part of speech, including verb in non-finite forms (infinitive, gerund, participle).

(1) \<NU\> Felicissima per il suo ritorno! \</NU\> (So happy for his return!)
(2) \<NU\> Il resto delle notizie su Solid Rock n. 267 \</NU\> (Other news in Solid Rock n. 267)
(3) \<NU\> Ma impegnarsi di più? \</NU\> (How about working harder?)
(4) \<NU\> Spariti i negozi, l'edicola, il posteggio. \</NU\> (Gone the shops, the newsstand, the parking space)
(5) \<NU\> Facendo due conti. \</NU\> (Doing the math)

NUs can have subordinate nominal clauses, also without a verb in a finite form. They should be included in the annotation.

(6) \<NU\> Che bello andare in vacanza tutti insieme! \</NU\> (So good to go to vacation all together!)
(7) \<NU\> Quanta poca umanità nell'aizzare l'opinione pubblica contro le minoranze per guadagnare più voti per le prossime elezioni, facendo il gioco della Russia. \</NU\> (How little humanity in inciting the public opinion against minorities in order to gain more votes in the next election, doing thus Russia's game.)

NUs can also have subordinate verbal clauses, with verbs in a finite form. They should be included in the annotation too.

(8) \<NU\> Una sofferenza che ricorda quella amorosa \</NU\> (A sufference that reminds me the one of love)
(9) \<NU\> Poche scuse, perché qui abbiamo finito! \</NU\> (No more excuses, because we're done here!)
(10) \<NU\> Poiché ora piove, tutti a casa! \</NU\> (Since it's raining, everybody home!)
(11) \<NU\> Veritiera l'opinione di Davide, che ha parlato ieri, sulla manovra finanziaria \</NU\> (A truthful opinion the one of Davide, who spoke yesterday, about the financial plan)

3. The annotation of coordinate NUs

3.1. Nominal + Verbal
In case a NU has a coordinate clause with a verb in a finite form, only annotate the NU, eventually with its subordinate clauses
(1) \<NU\> Poco male, \</NU\> non mi interessa. (No harm done, I don't care)
(2) \<NU\> Anche se siamo ferite, poco male, \</NU\> non mi interessa. (Even if we're hurt, no ham done, I don't care)

3.2. Nominal + Nominal

In case a NU has another NU as a coordinate clause, annotate both of them as a single NU. In fact, it is not always clear if we are looking at two distinct NUs or a the same NU with two subject with a common (implied) verb.

(1) <NU> Acqua a dirotto, tutti a casa! </NU> (Heavy rain, everybody home!)

(2) <NU> Giovedì sciopero, treni fermi otto ore. </NU> (Strike on Thrusday, trains stopped for eight hours)

(3) <NU> Sette regni, due continenti, mille personaggi, e tutti ad incontrarsi a caso. </NU> (Seven kingdoms, two continents, a thousand of characters, and everybody is meeting randomly.)

(4) <NU> Acqua a dirotto sul parcogiochi, che quindi chiude, e tutti a casa! </NU> (Heavy rain on the playground, which then closes, and everybody home!)

4. Greetings and courtesy formulae

Greeting and courtesy formulae without verbs in a finite form can occur as independent or coordinate NUs. Annotate them according to the previous paragraphs. Remember that the formulae like "*scusa/scusami*" are verbs in a finite form, and thus they should not be annotated as NUs.

(1) <NU> Ciao! </NU> (Hello!)

(2) <NU> Grazie! </NU> (Thank you)

(3) <NU> Ciao a tutti, </NU> è un piacere conoscervi (Hello everybody, it's a pleasure to meet you)

(4) <NU> Grazie per avermi aiutata, </NU> apprezzo molto (Thank you for helping me, I really appreciate it)

(5) <NU> Grazie per l'aiuto, un bacio! </NU> (Thank you for your help, a kiss!)

(6) <NU> Ciao! </NU> <NU> Piacere, Marta. </NU> (Hello! Greetings, Marta)

(7) <NU> Grazie per l'aiuto, </NU> e scusami per il disturbo! (Thanks for the help, and sorry for bothering you!)

5. Punctuation and limits of NUs

On Twitter, users can use punctuation in a non-standard way, thus it is necessary to set some rules for determining NUs' limits.

Generally, syntactic links take precedence over punctuation.

5.1. Strong punctuation

The full stop, the ellipsis, the exclamation point and the question mark are always NUs' delimiters, unless they clearly are separating words belonging to one clause, like in (5) and (6).

(1) <NU> Sì, proprio tu. </NU> Grazie per l'aiuto. </NU> Mi hai salvata. (Yes, right you. Thanks for the help. You saved me.)

(2) <NU> Ciao… </NU> sono nuova… <NU> piacere di conoscervi… </NU> (Hi… I'm new… nice to meet you…)

(3) <NU> Wow!!! </NU> <NU> Che bello! </NU> Sono felicissima! <NU> Grazie!!! </NU> (Wow!!! So nice! I'm so happy! Thank you!!!)

(4) <NU> Cosa??? </NU> <NU> Davvero?! </NU> Ma riesci a credergli? <NU> Seriamente??? </NU> (What??? Really?! Can you believe him? Seriously???)

(5) <NU> NIENTE. PIÙ. BIRRA! </NU> (NO. MORE. BEER!)

(6) <NU> Tanto rumore… per nulla! </NU> (So much noise… for nothing!)

(7) <NU> !!!Attenzione!!! </NU> (!!!Warning!!!)

(8) <NU> ???? </NU> Dici sul serio? (???? Are you serious?)

(9) <NU> Che ca….ta </NU> (Such bu….it)

5.2. Colon and semicolon

Usually, they are NUs' delimiters, unless they are used in a non-standard way and separate two sections of the same clause, like in (4) and (5).

(1) <NU> Conseguenze: </NU> <NU> innalzamento del livello dei mari </NU> (Consequences: raising of the sea level)

(2) <NU> Berlusconi: </NU> <NU> "Più lavoro per tutti!" </NU> (Berlusconi: "More jobs for everyone!")

(3) <NU> Casi studio: </NU> <NU> un cane con una rara malattia; </NU> <NU> un gatto acciecato a un occhio </NU> (Case study: a dog with a rare disease; a cat blind at one eye)

(4) Io porto: pizza e birra. (I'm going to bring: pizza and beer)

(5) <NU> Poco male; perché qui abbiamo finito </NU> (No harm done; because we're done here)

5.3. Punctuation in pairs: brackets, dashes and inverted commas

When brackets or dashes isolate an independent, non-defining non-verbal clause, consider them delimiters. Include the bracktes and the dashes inside the annotation when the NU correspond to the entire material inside the brackets/dashes, like in (1). Don't include the bracktes/dashes if the NU is only a part of the material inside the bracktes/dashes, as in (4).

(1) Il gioco non mi è piaciuto <NU> (nessuna sfida, troppi vantaggi, troppo oro, troppo veloce) </NU> (I didn't like the game (no challenges, too many advantages, too much gold, too much quick)

(2) <NU> Niente di cui preoccuparsi (per ora)! </NU> (Nothing to worry about (for now)!)

(3) <NU> Niente di cui preoccuparsi (visto che Lisa ha risolto il problema) </NU> (Nothing to worry about, since Lisa solved the problem)

(4) <NU> Buona notte! </NU> (<NU> Già sotto le coperte! </NU> Sono velocissima!) (Good night! (Already under the covers! I'm so quick!)

NB: the specification of a politician's political party between brackets is not an independent NU.

(5) <NU> Salvini (Lega): </NU> "Chiudiamo i porti!" (Salvini (Lega): "Let's close the harbors!"

When there is direct speech delimited by inverted commas or similar punctuation points, include the inverted commas inside the NU only if the NU correspond to the entire direct speech.

(6) Salvini dice: <NU> "Rimpatrio di tutti gli immigrati irregolari" </NU> (Salvini says: "Repatriation of all the illegal immigrants"

(7) Salvini dice: " <NU> Rimpatrio di tutti gli immigrati irregolari. </NU> Ci siamo stufati!" (Salvini says: "Repatriation of all the illegal immigrants. We're tired!")

5.4. Other graphic symbols: emoticons and dashes

When an emoticon is at the end of a NU, it should be included in the annotation, even if it is separated from the NU by a strong punctuation sign. If the emoticon stands between two independent utterances, it should be taken as a delimiter.

(1) <NU> Che bel vestito! *O* </NU> (Nice dress! *O*)

(2) <NU> Ciao a tutti 😊 </NU> (Hello everybody 😊)

(3) <NU> Ciao a tutti 😊 </NU> <NU> Che bel vestito! </NU> (Hello everybody 😊 Nice dress!)

Especially in tweets composed by newspapers' headlines, there could be one or more words separated from the rest of the tweet by a single dash. In this case, treat the dash as a strong punctuation sign.

(4) <NU> MILANO – </NU> <NU> Blitz dei carabinieri nel campo nomadi di Via Negrotto. </NU> (MILAN – Carabinieri's blitz in to the trailer park of Negrotto Street)

(5) <NU> Savona, problema sicurezza tra campo nomadi e profughi - </NU> <NU> Quotidiano online della provincia di Savona </NU> (Savona, security problem between trailer park and refugees – Online daily newspaper of Savona's province)

5.5. Absence of punctuation

Since Twitter's language is not a standard variety and has a character's limit, sometimes it is possible to have tweets without punctuation. In this case, analyze the tweet's syntactic structure and links, in order to understand if there is an independent NU: if a part of the tweet is clearly syntactically independent and without a verb in a finite form, annotate it as a NU.

(1) xché non facciamo un referendum abrogativo sul regio dec che obbliga a far entrate in Italia rom sinti ecc <NU> Affluenza alle stele </NU> (Why don't we do a referendum to recall the royal decree that force us to let in Italy Roma, Sinti ecc Sky-high turnout)

(2) <NU> maledetti immigrati </NU> gli italiani non hanno lavoro <NU> TUTTI A CASA! </NU> (Damn immigrants Italians don't have jobs EVERYBODY HOME!)

In some cases, there could be a string of coordinate NUs, or NUs syntactically independent from one another, but without punctuations signs to make it clear. Annotate them as one NU.

(3) Siamo stufi di pagare le tasse per questi immigrati <NU> rimpatriare tutti l'Italia agli italiani e non degli immigrati </NU> (We're tired of paying taxes for these immigrants repatriate everyone Italy to Italians and not to immigrants)

6. Twitter's peculiarities

The language on Twitter has some peculiar features that should be addressed. In fact, we should ask ourselves if and how hastags, at signs and links should be included inside NUs.

For this study on POP-HS-IT, we decided to focus on hashtags, signs and links well integrated inside the syntactic structure of the Tweet, excluding from the annotation those that are mere metadata.

6.1. Exclusion of metadata-like hashtags, at signs and links

If a hashtag, a Twitter's profile preceded by an at sign or a link are not integrated in the syntax of the comment, exclude them from the annotation of NUs. Note that these elements, theoretically, count as NUs, if they are non-verbal; still, for a study focused on the relationship between hate speech and NUs, these elements are not interesting, thus not relevant.

(1) #agorarai <NU> Col cavolo! </NU> (#agorarai Like hell!)

(2) <NU> Campi rom? </NU> <NU> Tutti da chiudere </NU> @matteosalvinimi @LegaNordPadania (Roma camps? They should be closed @matteosalvinimi @LegaNordPadania)

(3) Roma, il padre la maltratta: <NU> piccola bimba rom soccorsa dalla polizia </NU> https://t.co/2ygyijQElv (Rome, the father abuses her: young Roma child helped by the police https://t.co/2ygyijQElv)

6.2. Inclusion of syntactically integrated hashtags, at signs and links

If a hashtag, a Twitter's profile preceded by an at sign or a link are integrated in the syntax of the comment, include them in the NU's annotation.

(1) <NU> #Roma, avviato l'iter per il superamento dei campi #rom </NU> (#Rome, started the procedure for overcoming #Roma camps)

(2) <NU> Lavorare mai, @matteosalvini? </NU> (Never working, @matteosalvini?)

(3) <NU> Ecco le prove https://t.co/2gEv4bzVQ9 </NU> (Here is the evidence https://t.co/2gEv4bzVQ9)

6.3. Chopped sentences

The tweets on POP-HS-IT sometimes can be interrupted before a sentence is concluded, occasionally producing an unintended NU. These chopped sentences, if lacking a verb in a finite form, should be annotated as NUs.

(1) Rom borseggia un turista: <NU> I carabinieri della stazione… </NU> (Roma woman pickpockets a tourist: The Carabinieri (= police) at the station…)

(2) A mio parere l'unica cosa da fare era utilizzare i € 150.000,00 per lo smantellamento del campo nomadi. <NU> Campo tra... </NU> (In my opinion the only thing to do was using the € 150.000,00 to dismantle the Roma camp. Camp between…)

7. Other examples

(1) @Giustinianorex @ilmessaggeroit <NU> solo nomadi…TUTTI bianchi..neri.. </NU> e chi più ne ha più ne metta. Ci stanno DEVASTANDO siamo al limite! (@Giustinianorex @ilmessaggeroit only nomads…EVERYONE white..black..you name it. They're DEVASTATING us we are on the verge!)

(2) <NU> @DSantanche Sicura???? </NU> <NU> Anche per gli immigrati, i rom e i rifugiati? </NU> Chissà perché detto/scritto da te suona male (@DSantanche Sure???? For immigrants, Roma and refugees too? I don't know why, by told/written by you it sound suspicious)

(3) Rinchiudono 2 nomadi nel gabbiotto <NU> Denunciati tre addetti della Lidl (Corriere Adriatico) </NU> https://t.co/An6lra0St9 https://t.co/kMPWyvBPpu (Shut two nomads in a cage Reported three Lidl's workers (Adriatic Courier) https://t.co/An6lra0St9 https://t.co/kMPWyvBPpu)

(4) #Notizie #Ultimora <NU> Salvini: </NU> <NU> dopo i rom,via i centri sociali </NU> https://t.co/Y36h2gMF9S (#News #Latestevents Salvini: after Roma people, no more community centers https://t.co/Y36h2gMF9S)

(5) <NU> #Milano palme e cammelli </NU> ma a #Linate si cacciano ITALIANI senza tetto taglieggiati da rom. (#Milan palm trees and camels, but in #Linate homeless ITALIANS are kicked out extorting money from Roma people)

(6) <NU> Blitz della polizia al campo nomadi: </NU> <NU> trovate le borse appena rubate </NU> <NU> "PADOVA - </NU> <NU> Ampio controllo della Polizia... </NU> https://t.co/cmlcSKKlFZ (Police's blits at trailer park: found just stolen purses "PADUA – Wide Police's control… https://t.co/cmlcSKKlFZ)

(7) #Sala #Milano <NU> Ottimo lavoro! </NU> <NU> Citta sotto assedio. </NU> <NU> #Alfano: </NU> <NU> Stop profughi. </NU> <NU> necessario esercito x sicurezza cittadini </NU> (#Sala #Milan Great job! City under siege. #Alfano: Stop refugees. Army necessary for citizens' safety)

(8) @QuintaColonnaTv @GiorgiaMeloni <NU> a casa tutti quanti islamici musulmani clandestini e zingari! </NU> C'è poco da fare ! <NU> VIA TUTTI! </NU> (@FifhtColumnTv @GiorgiaMeloni all Islamics Muslims illegal immigrants and gypsies at home! There is little to do ! EVERYONE OUT!)

(9) @rcavalieri223 <NU> COR CAZZO!!! </NU> il radical shit ha un attico A NEW YORK MANHATTAN. <NU> Alla faccia degli immigrati, Africa, sinergie & cazz varie </NU> (@rcavalieri223 MY ASS!!! The radical shit has a penthouse IN NEW YORK MANHATTAN. Talking about immigrants, Africa, synergies & shit)

(10)   #Corona resta in carcere, <NU> i rom che rapinarono e provocarono la morte di #Zhang Yao impuniti. </NU> <NU> Che Italia triste. </NU> (#Corona stays in prison, the Roma who robbed and caused the death of #Zhang Yao unpunished. What a sad Italy.)