

# Guies de generació de parells de pregunta/resposta en català

Les pautes per escriure parells de preguntes i respostes per a cada paràgraf són les mateixes que SQuAD1.1 [[Rajpurkar et al., 2016](#)].

La tasca consisteix en presentar a l'anotador uns paràgrafs d'un article de la Viquipèdia catalana. Fent servir la informació presentada pels paràgrafs, s'han d'escriure entre una i cinc preguntes relacionades amb aquest contingut que puguin tenir una resposta a partir de citar TEXTUALMENT un fragment d'aquest paràgraf, sense afegir o elaborar res. La resposta només pot ser una cita textual (copy-paste) del paràgraf presentat.

## Exemple 1

### Paràgraf:

“Posteriorment, es va trobar un tros de paper en què Luter havia escrit la seva última declaració. La declaració era en llatí, a part de “Som captaires”, que era en alemany.”

### Pregunta 1:

“Què es va descobrir posteriorment escrit per Luter?”

### Resposta 1:

“la seva última declaració”

[En aquest cas una altra possible resposta seria simplement “un tros de paper”, però sempre triarem la més informativa, la que aporta més informació expressada en el paràgraf. Hem de fer *copy-paste* només del tros mínim que serveixi per respondre a la pregunta, i de cap manera fer edicions del tipus “un tros de paper [amb] la seva última declaració”]

### Pregunta 2 (sobre el mateix paràgraf):

“En quin idioma estava escrita la major part de la declaració?”

### Resposta 2:

“llatí”

[Podria haver-se triat “en llatí”, però el tros mínim informatiu no inclou la preposició]

## Exemple 2

«Al 1972, el físic francès Francis Perrin descobrí quinze reactors de fissió nuclear naturals antics i ja inactius en tres dipòsits de menes separats a la mina d'Oklo, al Gabon, a l'oest de l'Àfrica, que es coneixen col·lectivament com a «Reactors fòssils d'Oklo».»

### Pregunta:

On es va trobar Francis Perrin reactors de fissió nuclear?

### Resposta:

a la mina d'Oklo

La possible resposta “a la mina d'Oklo, al Gabon, a l'oest de l'Àfrica” aporta més informació, però no es rellevant ni central, i contesta més a una pregunta com «on és la mina d'Oklo?» que no pas la pregunta feta.

## Comentaris:

- És important que a la pregunta es pugui donar resposta només presentant un tros del text, sense necessitat de fer servir inferències addicionals o coneixement enciclopèdic de l'anotador. Les preguntes no han de ser massa llargues o confuses, ni fer servir recursos com la ironia, l'humor, etc.
- La possibilitat de generar entre una i cinc preguntes depèn de la informació rellevant continguda en els paràgrafs presentats. N'hi ha, potser, que només permeten fer una pregunta i altres que donen peu a més de cinc, però en farem un màxim cinc per paràgraf.
- Les preguntes no han de ser només un copy-paste de frases del paràgrafs, convertits a forma de pregunta. Han de ser formulacions en les seves pròpies paraules i directes, naturals d'un parlant de català.  
Així per exemple hem d'evitar preguntes com: “*Quan es va trobar un tros de paper en què Luter havia escrit la seva última declaració?*” on tota la pregunta és una còpia del text a excepció de l'interrogatiu ‘Quan’.
- També hem d'intentar que l'estil sigui directe i concís.  
Per exemple, enlloc de:  
«Als reactors trobats se'ls coneix també amb un altre nom, quin és aquest?»  
Millor:  
«Amb quin altre nom es coneixen els reactors trobats?»
- Les preguntes haurien de (tant com sigui possible) contemplar tota la varietat del ventall d'interrogacions possibles (*qui, què, com, on, quan*, etc.) donats els textos proporcionats, i per tant les seves “respostes” seran cites directes que continguin xifres, noms, dates, etc.
- Cal explicitar el referent a totes les preguntes sempre que sigui possible.  
Exemple a evitar:  
“En quin idioma estava escrita?”  
Exemple correcte:  
“En quin idioma estava escrita la seva última declaració?”  
És millor, però, evitar fer servir exactament les mateixes paraules que es presenten al paràgraf.  
Exemple millorat:  
“En quin idioma estava escrit el text?”
- Cal evitar fer preguntes sobre referents que puguin provocar ambigüitat o confusió, per exemple, si no són explícits al paràgraf mateix, encara que se n'hagi parlat en paràgrafs anteriors.