

# Article Title: document template for COLING-2012

*Annie SMITH*<sup>1,2</sup> *LI XiaoDong*<sup>1,3</sup>

(1) INSTITUTE\_1, address 1

(2) INSTITUTE\_2, address 2

(3) INSTITUTE\_3, address 3

Annie.Smith@institute.edu.ca, Li.Xiao-Dong@iitb.ac.in

## ABSTRACT

Place for the abstract in English (maximum 150 words).

This is a template for the submissions and proceedings of COLING-2012 and associated events. The Programme Committee and the Organizing Committee want to encourage participants to read as many contributions as possible during the conference and the workshops, and to lower the number of copies and the unit cost of paper proceedings. To ensure very good readability on PC screens as well as good readability of the paper in proceedings, authors are encouraged to use the *Charter* font (for LaTeX) and the *Charis* or *Georgia* font (for MS or Libre Office word processors) if they can, otherwise please use *Times New Roman*. Also, *the format will be A5* (2 pages on screen, 1 page on a larger standard format in print). The title of the article is in 14pt, bold, centered. Authors are in italics, 11pt, centered; affiliations in 9pt. E-mail addresses are in Courier New, 10pt, centered.

**AT SUBMISSION TIME, PLEASE ANONYMISE everywhere, REPLACING NAMES BY DUMMIES LIKE John DOE\_1, Jack DOE\_2.**

TITLE AND ABSTRACT IN ANOTHER LANGUAGE,  $L_2$  (OPTIONAL, AND ON SAME PAGE)

**Translation in  $L_2$  of the article title** (if option used):14pt, bold,centered

The translation of the title and abstract in another language  $L_2$  is optional. If the title is translated, the abstract must also be translated, as exactly as possible. In that case, put an abstract in  $L_2$  (maximum 170 words as translation expands text length) here.

Apart from enhancing visibility in one's own national scientific communities, these parallel abstracts might be the seed for a multilingual natural parallel corpus in computational linguistics. Tantus est igitur innatus in nobis cognitionis amor et scientiae ut nemo dubitare possit quin ad eas res hominum natura nullo emolumento invitata rapiatur. Videmusne ut pueri ne verberibus quidem a contemplantis rebus perquirendisque deterreantur ? Ut pulsi recurrant? Ut aliquid scire se gaudeant ? Ut id aliis narrare gestiant ? Ut pompa, ludis atque eius modi spectaculis teneantur ob eamque rem vel famem et sitim perferant ? Quid vero ?

---

**KEYWORDS:** Here a list of keywords in English.

**KEYWORDS IN  $L_2$ :** Here a list of keywords in  $L_2$  (if option used).

---

## 1 Optional condensed 2-page version in $L_2$ (1st section if present)

If the "other language" option is used on the first page, it is possible to put here a condensed version of the paper in the same  $L_2$  language. It should take no more than 2 pages.

The remaining of this section is not an example of a condensed version, rather a quick experiment to translate the above sentence with 2 widely known MT systems (PE below means PostEditing).

|    |         |  |
|----|---------|--|
| FR | Google  | Si le "autre langue" option est utilisée sur la première page, il est possible de mettre ici une version condensée du papier dans la même langue $L_2$ . Il ne devrait pas prendre plus de 2 pages.                  |
|    | Systran | Si l'option de la " autre langue " est employée à la première page, il est possible de mettre ici une version condensée du papier dans la même langue $L_2$ . Il devrait ne prendre pas plus de 2 pages.             |
|    | PE      | Si l'option " autre langue " est employée à la première page, il est possible de mettre ici une version condensée du papier dans la même langue $L_2$ . Cela devrait ne prendre pas plus de 2 pages.                 |
| DE | Google  | Wenn die "andere Sprache" Option auf der ersten Seite verwendet wird, ist es möglich, hier lege eine gekürzte Fassung des Papiers in der gleichen Sprache $L_2$ . Es sollte nicht mehr als 2 Seiten.                 |
|    | Systran | Wenn die "andere Sprach" Wahl auf der ersten Seite verwendet wird, ist es möglich, eine verkürzte Version des Papiers in die gleiche Sprache $L_2$ hier einzusetzen. Es sollte nicht mehr als 2 Seiten nehmen.       |
|    | PE      | Wenn die "andere Sprache" Wahl auf der ersten Seite verwendet wird, ist es möglich, eine verkürzte Version des Berichts in der gleichen Sprache $L_2$ hier einzusetzen. Es sollte nicht mehr als 2 Seiten einnehmen. |
| ZH | Google  | ZH content could be entered using native language support packages (if available)  |
|    | Systran | ZH content could be entered using native language support packages (if available)  |
|    | PE      | ZH content could be entered using native language support packages (if available)  |
| RU | Google  | RU content could be entered using native language support packages (if available)  |
|    | Systran | RU content could be entered using native language support packages (if available)  |
|    | PE      | RU content could be entered using native language support packages (if available)  |
| GR | Google  | GR content could be entered using native language support packages (if available)  |
|    | Systran | GR content could be entered using native language support packages (if available)  |
|    | PE      | GR content could be entered using native language support packages (if available)  |
| JP | Google  | JP content could be entered using native language support packages (if available)  |
|    | Systran | JP content could be entered using native language support packages (if available)  |

|    |         |  |
|----|---------|--|
|    | PE      | JP content could be entered using native language support packages (if available)  |
| AR | Google  | AR content could be entered using native language support packages (if available)  |
|    | Systran | AR content could be entered using native language support packages (if available)  |
| KR | Google  | KR content could be entered using native language support packages (if available)  |
|    | Systran | KR content could be entered using native language support packages (if available)  |
| ES | Google  | Si el "otro idioma" opción se utiliza en la primera página, es posible poner aquí una versión resumida del documento en el mismo lenguaje L <sub>2</sub> . Se debe tener no más de 2 páginas.                        |
|    | Systran | Si la opción de la "otra lengua" se utiliza en la primera página, es posible poner aquí una versión condensada del papel en la misma lengua L <sub>2</sub> . Debe tomar no más que 2 páginas.                        |
|    | PE      | Si la opción "otra lengua" se utiliza en la primera página, es posible poner aquí una versión condensada del papel en la misma lengua L <sub>2</sub> . Debe tomar no más que 2 páginas.                              |
| PR | Google  | Se o "outro idioma" opção é usada na primeira página, é possível colocar aqui uma versão condensada do papel na língua L <sub>2</sub> mesmo. Deve ter no máximo 2 páginas.   |
|    | Systran | Se da "a opção outra língua" é usada na primeira página, é possível pôr aqui uma versão condensada do papel na mesma língua L <sub>2</sub> . Deve tomar não mais de 2 páginas.                                       |
| IT | Google  | Se l' "altro linguaggio" opzione viene utilizzata nella prima pagina, è possibile mettere qui una versione condensata della carta nella stessa lingua L <sub>2</sub> . Essa non dovrebbe richiedere più di 2 pagine. |
|    | Systran | Se l' opzione dell' altra lingua" è usata alla prima pagina, è possibile mettere qui una versione condensata della carta nella stessa lingua L <sub>2</sub> . Dovrebbe prendere non più di 2 pagine.                 |
|    | PE      | Se l'opzione "altra lingua" è usata alla prima pagina, è possibile mettere qui una versione condensata del articolo nella stessa lingua L <sub>2</sub> . Dovrebbe prendere non più di 2 pagine.                      |
| HI | Google  | HI content could be entered using native language support packages (if available)  |
|    | Systran | HI content could be entered using native language support packages (if available)  |
|    | PE      | HI content could be entered using native language support packages (if available)  |

Urdu, Tamil and other languages don't display well at this moment (our fonts are not complete).

## 2 Main content of the paper (in English)

### 2.1 General information

This template continues as if the "other language" option had been chosen, meaning that the article proper, in English, begins on page 4. Its length corresponds to that of a short paper.

USUAL but IMPORTANT: papers must conform to the official COLING-2012 style guidelines (downloadable from the website).

Submission and reviewing will be on-line, managed by the START system. The only accepted format for submitted papers is PDF. Supplementary material, if any, in the form of tools or resources, must be in the form of a single .zip or a .tgz archive file with a maximum size of 10MB; otherwise there are no constraints on its format. Submissions, together with all supplementary material, must be uploaded on the START system by the submission deadlines; submissions after that time will not be reviewed.

To minimize network congestion, we request authors to upload their submissions as early as possible (especially if they contain large supplementary material files). Improved versions of the submissions may be uploaded until the deadline.

Submissions must be in the format described in this template. If for some reason you have a problem producing this format, please contact organisers as early as possible. Organisers reserve the right not to publish papers not conforming to the required format.

On acceptance of a submission, precise instructions will be given on how to send the camera ready version, and in some cases with required modifications. Camera ready versions not conforming to programme committee's improvement requests will be considered as unsent, and hence will not be presented at the conference and will not be included in the proceedings.

### 2.2 More on the format of submissions

Since several years, scientific conferences have been distributing proceedings on electronic media to avoid producing printed versions. Printed versions are given to only those participants who explicitly request and pay for them. Most participants get the proceedings on a CD or on a USB key. That is a very good idea, but the drawback is that very few participants manage to read more than a small number of papers, typically less than 1/10th or 1/20th of the number available in the proceedings. The main reason seems to be that it is very awkward to read traditional A4 sized portrait double-column articles on usual laptop screens.

The recent JEP-TALN-RECITAL-2012 conferences have successfully addressed that problem by proposing a template that is optimised for reading on PC screens. COLING-2012 and associated events propose to use it, with the option of using the Times New Roman font for authors having problems with installing the slightly better fonts Charter, Charis, or Georgia.

#### 2.2.1 Document format & character fonts

*Submissions will be in PDF, with page dimensions of 150 mm x 210 mm, and margins of 5 mm on all sides.* That corresponds to A5 (148.2 mm x 209.9 mm) with 4.1 mm horizontal margins and 5 mm vertical margins. To get the same layout on an A4 format while preparing a submission, choose A4 and put the margins at 35 mm (horizontal) and 48.3 mm (vertical).

*No header or footer should be present in submissions.*

The *preferred fonts* have been chosen because they are optimised for reading on screen while being elegant in print. If possible, submissions prepared with LaTeX should use the *Charter (Bitstream)* font. Users of Word or LibreOffice should use an equivalent font, namely (in order of preference):

1. *Charter* (via Bitstream), available on most Linux and in LaTeX *texlive* installations.
2. *Charis SIL* (via SIL). It is an extension of the Charter font, and is available at [http://scripts.sil.org/cms/scripts/page.php?item\\_id=CharisSIL\\_download](http://scripts.sil.org/cms/scripts/page.php?item_id=CharisSIL_download);
3. *Georgia*. This font is included in MS Word by default.

If authors encounter problems in installing one of the preferred fonts, they should use Times New Roman (available in all platforms) as default font, with the same point sizes.

*The body of the text should be in size 10pt, the titles in bold 11pt for levels 1 & 2, and 10pt from level 3 onwards.*

### 2.2.2 Pages and paragraphs

Paragraphs use single spacing and are justified left and right. They are not indented, but are separated by a half-line vertical space (6pt above and below).

The first page contains only the title, the names and affiliations of authors, the abstract(s) and the keywords (see page 1 of this template).

The body of the paper, including the introduction, begins on page 4 if a condensed (2-page) synopsis in another language  $L_2$  is included, and on page 2 otherwise.

A LaTeX stylesheet, a Word template, and an ODT template are available on the COLING-2012 website. Submissions will be handled interactively under SoftConf.

### 2.2.3 Submission sizes

Here are the length limits according to the type of submission. "Any number of reference pages" means that pages used for bibliographical and Internet references are not counted in the page number limits. However, it stands to reason that reviewers will accept only reasonable sizes.

|              |  |                     |
|--------------|--|---------------------|
| full papers  | 14 A5 pages with any number of reference pages | Oral presentation   |
| short papers | 8 A5 pages with any number of reference pages  | Poster presentation |
| demo papers  | 6 A5 pages with any number of reference pages  | Demonstration       |

Papers are in English. However, if an author so wishes s/he may add elements in another language ( $L_2$ ): a title, an abstract and keywords on the first page, and 2 pages (a synopsis) in  $L_2$ , beginning on page 2. Such papers will have 14+2 (or 8+2, or 6+2) pages, plus any number of pages for reference.

## 2.3 Other formatting elements

### 2.3.1 Lists

Here is a bulleted list, where each element begins with an independent sentence (capitalized).

- Its lines are separated by less vertical space than normal paragraphs.
- We don't add space between paragraphs of that same style.

Here is a numbered list of items that are parts of an outside sentence. Take care to:

1. restart the numbering and get the numbers consecutive.
2. not to capitalize the initial letter of such items.

### 2.3.2 Figures and tables

Figures and tables will be centered on the page, with a caption underneath. The caption will contain the keyword `FIGURE` or `TABLE` in small capitals, followed by the number of the figure or table. Numbers in each category are independent. Examples appear in this template. Equations and formulae may appear in line or centered on the page, without captions. A reference number may appear to the right of a formula or an equation.

|         |   |
|---------|---|
| A table |   |
|         | The table is centered. Its cells may be optionally centered. Avoid full justification leading to very long horizontal spaces. |

Table 1: A table caption appears under it.

### 2.3.3 Notes and references

Footnotes<sup>1</sup> can be used.

References to the bibliography, for example (Bernhard, 2007), are formatted with the APA style (see <http://www.apastyle.org/>). It is available in all the LaTeX installations based on *texlive* and on CTAN.

Users of Word or LibreOffice are encouraged to follow this style as much as possible when using their bibliography management tool (Mendeley, Zotero or Endnote). They can refer to the example below. Authors citing online references (URLs) should include the creation date in the reference.

The rest of this document only illustrates the format described above.

Tantus est igitur innatus in nobis cognitionis amor et scientiae ut nemo dubitare possit quin ad eas res hominum natura nullo emolumento invitata rapiatur. Videmusne ut pueri ne verberibus quidem a contemplandis rebus perquirendisque deterreantur ? Ut pulsi recurrant? Ut aliquid scire se gaudeant ? Ut id aliis narrare gestiant ? Ut pompa, ludis atque eius modi spectaculis teneantur ob eamque rem vel famem et sitim perferant ? Quid vero ? Qui ingeniis studiis atque artibus delectantur, nonne videmus eos nec valetudinis nec rei familiaris habere rationem omniaque perpeti ipsa cognitione et scientia captos et cum maximis curis et laboribus compensare eam quam ex discendo capiant voluptatem ?

---

<sup>1</sup>Here is one!



Figure 1: An image used as a figure: It is a good idea to put each figure in a table cell to avoid (in Word) problems with anchored or floating figures.

Omne hoc tempus inter pugillares ac libellos iucundissima quiete transmisi. "Quem ad modum, inquis, in urbe potuisti ?"

Circenses erant, quo genere spectaculi ne levissime quidem teneor. nihil novum, nihil varium, nihil quod non semel spectasse sufficiat. Quo magis miror tot milia virorum tam pueriliter identidem cupere currentis equos, insistentis curribus homines videre.

Si tamen aut velocitate equorum aut hominum arte traherentur, esset ratio non nulla; nunc favent panno, pannum amant, et, si in ipso cursu medioque certamine hic color illuc, ille huc transferatur, studium favorque transibit, et repente agitadores illos, equos illos, quos procul noscitant, quorum clamitant nomina, relinquunt. Tanta gratia, tanta auctoritas in una vilissima tunica, mitto apud volgus, quod vilius tunica, sed apud quosdam graves homines; quos ego cum recordor in re inani, frigida, assidua tam insatiabiliter desiderare, capio aliquam voluptatem, quod hac voluptate non capior.

Ac per hos dies libentissime otium meum in litteris colloco, quos alii otiosissimis occupationibus perdunt. Vale. (PLINE Junior, Letters, IX, 6.)

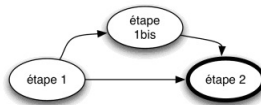


Figure 2: It is not advised to use Word directly to generate drawings. Use another tool (even Powerpoint) and insert the resulting image (at least 300 DPI or in vectorised format).

### 3 Title of a section

Rem atrocem nec tantum epistola dignam Larcus Macedo, vir praetorius, a servis suis passus est, superbus alioqui dominus et saevus et qui servisse patrem suum parum, immo nimium meminisset. Lavabatur in villa Formiana; repente eum servi circumstant, alius fauces invadit, alius os verberat, alius pectus et ventrem atque etiam, foedum dictu, verenda contundit; et, cum exanimem putarent, abiciunt in fervens pavimentum, ut experirentur an viveret.

Ille, sive quia non sentiebat, sive quia non sentire se simulabat, immobilis et extensus fidem peractae mortis implevit. Tum demum quasi aestu solutus effertur; excipiunt servi fideiores, concubinae cum ululatu et clamore concurrunt. Ita et vocibus excitatus et recreatus loci frigore, sublatis oculis agitatoque corpore vivere se - et iam tutum erat - confitetur.

Diffugiunt servi quorum magna pars comprehensa est, ceteri requiruntur; ipse paucis diebus aegre focilatus non sine ultionis solacio decessit, ita vivus vindicatus ut occisi solent. (PLINE, Letters, III, 14.)

### 3.1 Title of a subsection

Miles cum oboediens potestati, sub qua legitime constitutus est, hominem occidit, nulla civitatis suae lege reus est homicidii; immo, nisi fecerit, reus est imperii deserti atque contempti. Quod si sua sponte atque auctoritate fecisset, in crimen effusi humani sanguinis incidisset. Itaque unde punitur si fecit iniussus, inde punitur nisi fecerit iussus. Quod si ita est, iubente imperatore, quanto magis, iubente creatore !

#### 3.1.1 Title of a subsection

Qui ergo audit non licere se occidere, faciat, si iussit cuius non licet iussa contemnere; tantummodo videat utrum divina iussio nullo nutet incerto... Hoc dicimus, hoc asserimus, hoc modis omnibus approbamus neminem spontaneam mortem sibi inferre debere velut fugiendo molestias temporales, ne incidat in perpetuas. (Sanctus Augustinus.)

### Conclusion and perspectives (not numbered, use style “Heading1”)

This section should better not be numbered (style “Heading1” inherits from “Heading 1”).

### Acknowledgments (not numbered, use style Heading1)

Optional section. If some named entities may lead to authors identification, please anonymize them.

Thanks to Gilles Sérasset for having kindly shared with us the templates he devised for TALN-JEP-RECITAL-2012.

### References

Aho, A. V. and Ullman, J. (1972). *The Theory of Parsing, Translation and Compiling*, volume 1. Prentice-Hall, EngleWood Cliffs, NJ.

Ashok K. Chandra, D. C. K. and J.Stockmeyer, L. (1981). Alternation. *Journal of the Association for Computing Machinery*, 28(1):114–133.

Benarmara, F., Hatout, N., Muller, P., and Ozdowska, S., editors (2007). *Actes de TALN 2007 (Traitement automatique des langues naturelles)*, Toulouse. ATALA, IRIT.

Bernhard, D. (2007). Apprentissage non supervisé de familles morphologiques par classification ascendante hiérarchique. In (Benarmara et al., 2007), pages 367–376.

Laignelet, M. and Rioult, F. (2009). Repérer automatiquement les segments obsolescents à l’aide d’indices sémantiques et discursifs. In *Actes de TALN 2009 (Traitement automatique des langues naturelles)*, Senlis. ATALA, LIPN.

Langlais, P. and Patry, A. (2007). Enrichissement d’un lexique bilingue par analogie. In (Benarmara et al., 2007), pages 101–110.

Mukherjee, S. and Bhattacharyya, P. (2012). Wikisent : Weakly supervised sentiment analysis through extractive summarization with wikipedia. In *European Conference on Machine Learning (ECML-PKDD 2012)*, pages 250–260, Bristol, U.K. Springer.