

Journal of Language Modelling

Adam Przepiórkowski

Institute of Computer Science, Polish Academy of Sciences,
Warsaw, Poland

1

YET ANOTHER JOURNAL?

Welcome to the inaugural issue of the Journal of Language Modelling (JLM), a free open-access peer-reviewed journal aiming to help bridge the gap between theoretical linguistics and natural language processing (NLP).

Setting up a new journal is not a trivial task, and running it possibly for decades requires determination and perseverance, so any such enterprise should not be taken up lightly. The publication of this issue has been preceded by years of growing conviction that there is no appropriate forum for the exchange of ideas between theoretical, formal and computational linguists. Many conversations with our colleagues – both linguists and NLP practitioners – convinced us that such a forum is indeed needed.

Ideally, JLM papers should be accessible to many readers of such periodicals as *Natural Language and Linguistic Theories*, *Journal of Linguistics*, *Language* or *Lingua* on one hand, and *Computational Linguistics*, *Journal of Natural Language Processing*, *Journal of Logic, Language and Information* or *Language Resources and Evaluation*, on the other. The affinity to another relatively young journal, *Linguistic Issues in Language Technology*, should also be clear. On the map of the main linguistic and NLP conferences, we see JLM as close to conferences devoted to constraint-based and formal linguistic theories (HPSG, LFG, TAG, Construction Grammar; Dependency Grammar in general and Meaning-Text Theory in particular; etc.), the *Formal Grammar* conference at ESSLLI, *COLING*, *Treebanks and Linguistic Theories*, etc., but also to *LREC (Language Resources and Evaluation Conference)*, *TSD (Text, Speech and Dialogue)* or the *xTAL* series of conferences (see *Jap-*

TAL 2012, *IceTAL 2010*, *GoTAL 2008*, etc.) on one hand, and *WCCFL* (*West Coast Conference on Formal Linguistics*) or *CSSP* (*Colloque de Syntaxe et Sémantique à Paris*) on the other. The connection between JLM and conferences which concentrate on statistical methods and numerical evaluation, such as *ACL* and *CICLing*, is more complex: results presented at such conferences are relevant to JLM to the extent they say something new about language, not just about usability in an NLP task, and to the extent they are accessible to potentially interested linguists.

It should be clear, thus, that we understand the discipline of language modelling very broadly – much more broadly than normally construed in speech recognition or statistical machine translation. Typical articles published in JLM are expected to deal with linguistic generalisations – their application in natural language processing and their discovery in language corpora – but possible topics range from precise linguistic analyses of phonological, morphological, syntactic, semantic and pragmatic language phenomena to mathematical models of aspects of language, and further to computational systems making non-trivial use of linguistic insights.

2

HOW IT WORKS

JLM is free for everybody – readers and authors alike. All journal content appears on a Creative Commons licence. We strongly recommend the Creative Commons Attribution 3.0 Unported licence (<http://creativecommons.org/licenses/by/3.0/>), which is the default licence for JLM articles,¹ but currently authors have the option to choose a different CC licence.

JLM is published electronically at <http://jlm.ipipan.waw.pl/>, with a printing on demand option (at a fee) coming soon. Once this inaugural issue is published, the journal will receive an ISSN number and every necessary effort will be made to include JLM in all relevant databases and indexes.

Papers are reviewed within less than three months of their receipt (with every effort made to minimise the reviewing period, while

¹ See Stuart Shieber's position paper in this issue on why pure CC-BY should be employed.

maintaining the quality), and they appear as soon as they have been accepted – there are no delays typical of traditional paper journals. Accepted articles are then collected in half-yearly numbers and yearly volumes, with continuous page numbering.

Journal of Language Modelling has a fully traditional view of quality: all papers are carefully refereed by at least three reviewers (including at least one member of the Editorial Board) and they are only accepted if they adhere to the highest scientific, typographic and stylistic standards. It is a contentious issue whether double-blind reviewing is indeed the best reviewing model,² but it is suggested by indexing agencies, so we adopt it for the time being. A list of reviewers will be published annually or bi-annually, depending on the turnout (to preserve anonymity).

The current make-up of the Editorial Board (EB) is available at the JLM website. The EB will partially rotate every couple of years, depending mainly on the profile of papers actually submitted to JLM. Day-to-day management of the journal will be performed by selected members of the Linguistic Engineering Group at the Institute of Computer Science, Polish Academy of Sciences (<http://zil.ipipan.waw.pl/>), which hosts the journal. Currently, the Managing Editors are Elżbieta Hajnicz, Agnieszka Mykowiecka, Adam Przepiórkowski and Marcin Woliński. The layout of JLM has been designed by Adam Twardoch and the corresponding L^AT_EX class has been developed by Marcin Woliński, who is also the main Layout Editor. Łukasz Dębowski acts as the Statistical Editor and rotating PhD students act as Copy Editors. All work at JLM is currently performed on voluntary basis and we are counting on community support in order to maintain this business model.³

3

THIS ISSUE

This inaugural issue is not a typical JLM issue. It contains two main sections: EDITORIALS and ARTICLES. The former, EDITORIALS, con-

²See, e.g., <http://www.linguateca.pt/Diana/SignedReviews.html> for arguments for giving up any anonymity.

³Offers from PhD students in Linguistics or Natural Language Processing – English native speakers, to act as Copy Editors, and L^AT_EX or X_YL^AT_EX specialists, to help with Layout Editing – are particularly welcome.

sists of three papers: the current editorial, Stuart Shieber's position paper on Creative Commons and Stefan Müller's position paper on Open Access in Linguistics. This way JLM joins the discussion about publishing models in science and makes a clear statement in support of maximal openness.

This issue's ARTICLES section, on the other hand, consists of four papers by partners of CESAR, a European project which sponsors the launch of the *Journal of Language Modelling*.⁴ The first paper, by Radovan Garabík and Mária Šimková, presents the morphosyntactic tagset used in the annotation of the Slovak National Corpus. The next paper, by Svetla Koeva et al., describes design considerations behind the Bulgarian National Corpus. The third paper, by Krešimir Šojat, Matea Srebačić and Marko Tadić, deals with morphosemantic relations between verbs in the Croatian WordNet. The final paper, by György Szaszák and András Beke, is concerned with the the prosody-to-syntax mapping in Hungarian. Note that the languages considered in these papers reflect the linguistic scope of the CESAR project and, similarly, the range of topics reflects the main theme of the project, i.e., language resources. While these articles (and other papers submitted to this issue) underwent the full reviewing procedure, this selection should not be taken as fully representative of JLM, which – as indicated above – has a much broader scope.

4

INSTEAD OF CONCLUSION

Launching a new journal is a high-risk business. Many people have already invested much of their time in this initiative, and returns are far from certain – the community may or may not accept it. We, the Managing Editors, are cautiously optimistic; responses to invitations to join the Editorial Board have overall been very positive, and 9 articles are currently under review for the first regular issue of JLM, to appear in mid-2013. We hope that the often enthusiastic reaction of our colleagues to the idea of JLM is shared by a sufficient number of theoretical, formal and computational linguists to make the current enterprise viable.

⁴CESAR stands for *CENTRAL and South-east EuropeAN Resources*; it is a part of the META-NET initiative. See <http://www.cesar-project.net/>.