**BAD: An assistant tool for making verses in Basque**

**Manex Agirrezabal, In˜ aki Alegria, Bertol Arrieta**

University of the Basque Country (UPV/EHU)

[maguirrezaba008@ikasle.ehu.es,](mailto:maguirrezaba008@ikasle.ehu.es) [i.alegria@ehu.es,](mailto:i.alegria@ehu.es) [bertol@ehu.es](mailto:bertol@ehu.es)

**Mans Hulden**

Ikerbasque (Basque Science Foundation)

[mhulden@email.arizona.edu](mailto:mhulden@email.arizona.edu)

**Abstract**

We present work on a verse-composition assistant for composing, checking correct- ness of, and singing traditional Basque *bertsoak*—impromptu verses on particular themes. A performing *bertsolari*—a verse singer in the Basque Country—must ad- here to strict rules that dictate the format and content of the verses sung. To help the aspiring *bertsolari*, we provide a tool that includes a web interface that is able to analyze, correct, provide suggestions and synonyms, and tentatively also sing (using text-to-speech synthesis) verses composed by the user.

**1 Introduction**

In the Basque Country there exists a long- standing live performance tradition of improvis- ing verses—a type of ex tempore composition and singing called *bertsolaritza*. Verses in *bertsolar- itza* can be seen as discourses with strict rules governing the technical structure of them: verses must contain a certain number of lines and each line must have a defined number of syllables, cer- tain lines have to rhyme in certain patterns, and so forth.

In this paper we present a web-based assistant tool for constructing verses (*bertsoak*) according to the rules of *bertsolaritza* (Garzia et al, 2001).

If the reader is interested in this topic, we rec- ommend watching the 2011 film *Bertsolari*1 2, di- rected by Asier Altuna.

1 IMDB: <http://www.imdb.com/title/tt2058583>

2 Trailer on: <http://vimeo.com/9355066>

**2 Relationship to earlier work**

There exist some prior works dealing with Basque verse-making and computer technologies, such as BertsolariXa (Arrieta et al., 2001), which is a rhyme search tool implemented as finite-state au- tomata using the two-level morphology formal- ism. The tool also contains other features, includ- ing semantic categorization of words, narrowing word-searches to certain themes, etc. While Bert- solariXa focuses mostly on the word-level, the current work also includes constraints on over- all verse structure in its implementation as well as a synonym search tool, a melody suggestion system, and possibilities for plugging in text-to- speech synthesis of verses.

**2.1 The *Bertsolari* tradition**

*Bertsolaritza* is very ingrained in the Basque Country and championships, competitions and get-togethers on *bertsolaritza* are quite common. Usually the competitors in such event, called *bert- solaris*, are given a theme to produce a verse on under some very limited time constraints.

But the Basque Country is not the only place that hosts such troubadour traditions—similar customs are present in many other countries such as Cuba, Brazil, Argentina, etc. The goal of the current tool is to be generalizable, and so appli- cable to various strategies of verse improvisation, and possibly be useful not only for Basque speak- ers, but also for others.

Below we briefly present an example of a verse made in the Basque Country. In 1986 Andoni Egan˜ a (a well-known *bertsolari*) was asked to sing a *bertso* and assigned a topic. In the verse, he was asked to play the role of an old person who lived alone, and who realized that he could

13

*Proceedings of the 6th EACL Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 13–17, Avignon, France, 24 April 2012. *Q*c 2012 Association for Computational Linguistics

not even tie his shoes. Within a few seconds he composed and sang three verses. Here, we ana- lyze the first verse.

Verse:

Gazte aroan ibili arren gustora tirriki-tarra,

denbora honen joan etorriak ederki jo dit gitarra,

gorputza daukat ximeldurikan ta eskuen punta zaharra, denborarekin seko galdu det gazte aroko indarra,

ez al da pena gizon mardul bat hola ibili beharra.

Translation:

Even when I was young I was always on a spree over time

I have been punished

I have a crumpled body

and the tip of the hands very old, Over time I lost

the strength I had when I was young, It’s a shame that a strong man

has to end up like me.

The special charm of *bertsolaritza* improvi- sation is that people proficient in the art can quickly express a variety of ideas, although they are working with very restrictive rules concern- ing the number of syllables in words they use, and how the words must rhyme. We must take into account that Andoni Egan˜ a was able to sing this verse within a few seconds of being given the topic, and also, that it complies exactly with a certain metric. In this case, the verse contains eight lines, each odd line consisting of ten sylla- bles, and each even line of eight syllables, with the even lines rhyming.

Formal training in the *bertsolari* tradition also exists in the Basque Country. In the last 20 to

30 years, an important movement has developed that aims to provide instruction to upcoming gen- erations on how to create verses (orally or in writing). This kind of instruction usually takes place in learning centers called *bertso-eskolak*, which in English roughly means, “verse-making schools.” The proliferation of this movement has produced a strong base of young *bertsolaris*, of whom many achieve an outstanding level of im- provisation skills.

**3 The BAD tool**

BAD is the acronym for “*Bertsotarako Arbel Dig- itala*”, roughly “Digital verse board.” The aim of the tool is to serve as a general assistant for *bertsolari*-style verse composition and help verse- making learners in their learning process.

This tool has been developed using the PHP programming language, but it contains certain parts developed using finite-state technology. The main functions of this tool, which will be dis- cussed in more detail in the next five sections, are the following: visualization of the verse structure, structure checking, rhyme and synonym searching and verse singing.

**3.1 Verse structure**

The main rules of the *bertsolari* verse are that a verse must consist of a certain predefined number of lines and each line in turn, of a predefined num- ber of syllables. Traditionally, about a hundred different schemes are used, and the tool provides support for all these patterns. For example, the structure called “*Hamarreko handia*” has ten lines and ten syllables in the odd-numbered lines, and eight syllables in the even-numbered lines. In this structure, the even-numbered lines have to rhyme. Selecting this scheme, the tool will mark the cor- responding lines with their requirements.

The web interface can be seen in figure 1, which shows the general layout of the tool, illus- trated with the example verse referred to above— we see that each line has been approved in terms of line length and syllable structure by the tool.

We have designed a database in which the main verse structures are saved so that when the user se- lects one verse schema, the system knows exactly the number of lines it must contain, where must it rhyme and how many syllables each line should have. Those schemata are also linked to melodies, each melody corresponding to one possible struc- ture.

**3.2 Structure checking**

After writing the verse, the system can evaluate if it is technically correct, i.e. if the overall structure is correct and if each line in the form abides by the required syllable count and rhyming scheme. The syllable counter is implemented using the *foma* software (Hulden, 2009), and the implementation (Hulden, 2006) can be found on the homepage of

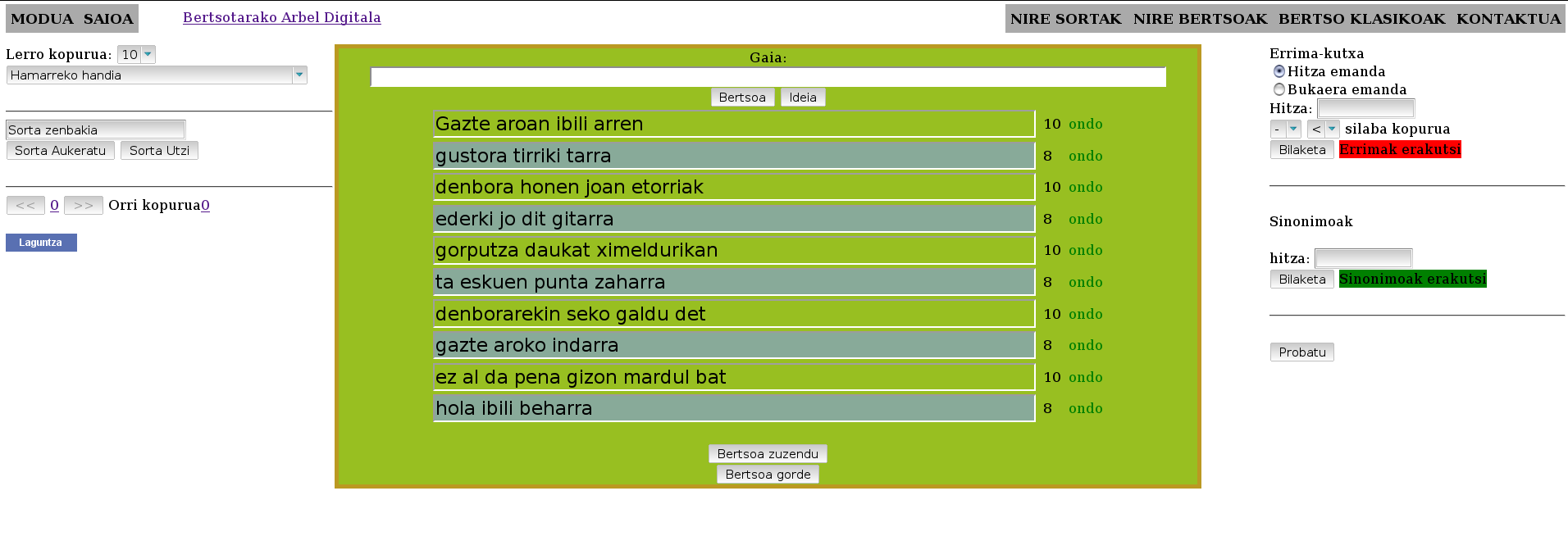


Figure 1: A verse written in the BAD web application.

*foma*.3

Separately, we have also developed a rhyme checker, which extracts special patterns in the lines that must rhyme and checks their confor- mity.

These patterns are extracted using *foma* (see section 3.4) after which some phonological rules are applied. For example, an example rule *era*

*→ {*era, eda, ega, eba*}*, models the fact that any word ending in *era*, for example, *etxera*, will rhyme with all words that end in *era*, *eda*, *eba* or *ega*. These rhyming patterns have been extracted according to the phonological laws described in (Amuriza, 1981).

**3.3 Synonym search**

Usually, people who write verses tend to quickly exhaust their vocabulary and ideas with to ex- press what they want to say, or encounter prob- lems with the number of syllables in various ten- tative words they have in mind. For example, if the verse-maker wants to say something con- taining the word “family,” (*familia* in Euskera, a four-syllable word) but is forced to use a three- syllable word in a particular context, the inter- face provides for possibilities to look for three- syllable synonyms of the word *familia*, producing the word *sendia*— a word whose meaning is oth- erwise the same, and made up of three syllables.

For developing the synonym search, we used a modified version of the Basque Wordnet (Pociello

3 [http://foma.googlecode.com](http://foma.googlecode.com/)

et al., 2010), originally developed by the IXA group at the University of the Basque Country. Within Wordnet we search the synsets for the in- coming word, and the words that correspond to those synsets are returned.

**3.4 Rhyme search**

The classical and most well-known problem in *bertsolaritza* concern the rhyming patterns. As mentioned, various lines within a verse are re- quired to rhyme, according to certain predefined schemata. To search for words that rhyme with other words in a verse, the BAD tool contains a rhyme search engine. In the interface, this is lo- cated in the right part of the BAD tool main view, as seen in figure 2.

The rhyme searcher is built upon finite-state technology, commonly used for developing mor- phological and phonological analyzers, and calls upon the freely available *foma*-tool, to calculate matching and nonmatching rhyme schemes.

Its grammar is made up of regular expressions that are used to identify phonological patterns in final syllables in the input word. The result of the search is the intersection of these patterns and all the words generated from a morphological de- scription of Basque (Alegria et al., 1996)—that is, a list of all words that match both the required phonological constraints given (rhyming) and a morphological description of Basque.

Based upon figure 2, if we search rhymes for the word *landa* (cottage), the system proposes a



Figure 2: The response of the rhyme search engine.

set of words that can be filtered depending on the number of syllables required. Among this list of words, we can find some words that end in *anda*, such as, *Irlanda* (Ireland) or *eztanda* (explosion), but through the application of phonological equiv- alency rules we also find terms like *ganga* (vault).

**3.5 Singing synthesis**

Another characteristic, as mentioned, is that, in the end, the verses are intended to be sung in- stead of only being textually represented. Based on other ongoing work in singing synthesis, we have designed a system for singing the verses en- tered into the system in Basque.

This is based on the “singing mode” of the Fes- tival text-to-speech system (Taylor et al., 1998). The advantage of using this is that Festival is open-source and has given us ample opportunities to modify its behavior. However, as Festival does not currently support Basque directly, we have re- lied on the Spanish support of the Festival sys- tem.4

4 While morphologically and syntactically, Spanish and Basque have no relationship whatsoever, phonetically the languages are quite close, with only a few phonemes, syl-

Based on current work by the Aholab research team in Bilbao—a lab that works on Basque speech synthesis and recognition—we have im- plemented a singing module for BAD, based on the text-to-speech HTS engine (Erro et al., 2010). Our application is able to sing the composed verses entered into the system in Basque, with a choice of various standard melodies for *bertsolar- itza*.5

**4 Discussion and future work**

Now that the BAD tool has been developed, our intention is to evaluate it. To make a qualita- tive evaluation we have gotten in touch with some verse-making schools (*bertso-eskola*), so that they can test the system and send us their feedback us- ing a form. Once the evaluation is made, we will improve it according to the feedback and the sys- tem will be made public.

Our ultimate goal is to develop a system able to create verses automatically. To achieve this long- term goal, there is plenty of work to do and ba- sic research to be done. We have in our hands a good corpus of 3,500 Basque verse transcriptions, so we intend to study these verses from a mor- phological, syntactical, semantical and pragmatic point of view.

In the short term, we also plan to expand the synonym search to be able to provide searches for semantically related words and subjects (and not just synonyms), like hypernyms or hyponyms. The Basque WordNet provides a good opportu- nity for this, as one is easily able to traverse the WordNet to encounter words with varying degrees of semantic similarity.

Another feature that we want to develop is a system that receives as input a verse together with a MIDI file, and where the system automatically sings the verse to the music provided.

Finally, in order for the system to be able to provide better proposals for the verse artist—including perhaps humorous and creative proposals—we intend to work with approaches to computational creativity. We are considering different approaches to this topic, such as in the work on *Hahacronym* (Stock et al., 2005) or the *Standup* riddle builder (Ritchie et al., 2001).

labification rules, and stress rules being different enough to disturb the system’s behavior.

5 However, this functionality is not available on the web

interface as of yet.

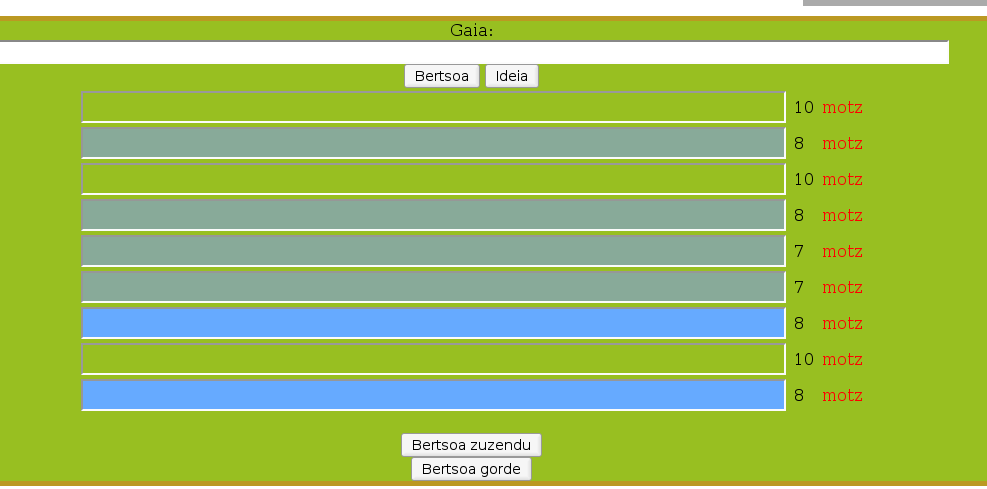


Figure 3: The BAD application before entering a verse, showing two possible rhyme patterns.

**Acknowledgments**

This research has been partially funded by the

Basque Government (Research Groups, IT344-

10).

**References**

In˜ aki Alegria, Xabier Artola, Kepa Sarasola and Miriam Urkia, “Automatic morphological analysis of Basque”, Literary and Linguistic Computing, ALLC, 1996

Xabier Amuriza, “Hiztegi errimatua”, Alfabetatze Eu- skalduntze Koordinakundea, 1981

Bertol Arrieta, In˜ aki Alegria, Xabier Arregi, “An as- sistant tool for Verse-Making in Basque based on Two-Level Morphology”, 2001

Daniel Erro, In˜ aki Sainz, Ibon Saratxaga, Eva Navas, Inma Herna´ez, “MFCC+F0 Extraction and Wave- form Reconstruction using HNM: Preliminary Re- sults in an HMM-based Synthesizer”, 2010

Joxerra Garzia, Jon Sarasua, and Andoni Egan˜ a, “The art of bertsolaritza: improvised Basque verse singing”, Bertsolari liburuak, 2001

John E. Hopcroft, Rajeev Motwani, Jeffrey D. Ullman, “Introduccio´ n a la teor´ıa de Auto´ matas, Lenguajes y Computacio´ n”, Pearson educacio´ n, 2002

Mans Hulden, “Foma: a finite-state compiler and li- brary”, Proceedings of the 12th Conference of the European Chapter of the Association for Computa- tional Linguistics: Demonstrations Session, p. 29–

32, 2009

Mans Hulden, “Finite-state syllabification”, Finite- State Methods and Natural Language Processing, p.

86 – 96, Springer, 2006

Kimmo Koskenniemi. “Two-level morphology: A general computational model for word-form recog- nition and production”. Publication 11, University of Helsinki, Department of General Linguistics, Helsinki, 1983.

Eli Pociello, Eneko Agirre, Izaskun Aldezabal, “Methodology and construction of the Basque WordNet”, 2010

Graeme Ritchie, “Current directions in computational humour”, Artificial Intelligence Review, Volume

16, Number 2, p. 119 – 135, Springer, 2001

Graeme Ritchie, Ruli Manurung, Helen Pain, Annalu Waller, Dave O’Mara, “The STANDUP interactive riddle builder”, Volume 2, Number 2, p. 67 – 69, IEEE Intelligent Systems, 2006

Oliviero Stock and Carlo Strapparava, “Hahacronym: A computational humor system”, Proceedings of the ACL 2005 on Interactive poster and demonstra- tion sessions, p. 113 – 116, Association for Compu- tational Linguistics, 2005

Paul Taylor, Alan W. Black and Richard Caley, “The architecture of the Festival speech synthesis sys- tem”, International Speech Communication Asso- ciation, 1998