Hand-out Th. Fransen - CLC10 05/09/18

Trinity College Dublin
fransent@tcd.ie

ILLUSTRATING FINITE-STATE TRANSDUCERS FOR OLD IRISH VERBS

(**L** = lexical level, **S** = surface level)

(1) Ambiguity with morphological (= word-based) analysis: 2 x 8 = 16 lexical / upper-level combinations for surface input *na␣mbrissi* (separated by space), with the three correct/possible analysis combinations in bold.

	Proclitic element	Stem and ending (+ nota aug.)
L (1)	**no+CONJ_PART+PRON+3P+SG+MASC+NAS**	DEPEND+SIMPLE+NAS+bris+V +W2a+IMPV+CONJ+2P+SG ~~+NOTA_AUG+2P+PL~~
L (2)	no+CONJ_PART+PRON+3P+SG~~+NEUT+LEN~~	DEPEND+SIMPLE+NAS+bris+V +W2a+IMPV+CONJ+2P+SG ~~+NOTA_AUG+3P+SG+FEM~~
L (3)		**DEPEND+SIMPLE+NAS+bris+V +W2a+PRES+IND+CONJ+2P+SG**
L (4)		**DEPEND+SIMPLE+NAS+bris+V +W2a+PRES+IND+CONJ+3P+SG**
L (5)		DEPEND+SIMPLE+NAS+bris+V +W2a+PRES+SUBJ+CONJ+1P+SG ~~+NOTA_AUG+2P+PL~~
L (6)		DEPEND+SIMPLE+NAS+bris+V +W2a+PRES+SUBJ+CONJ+1P+SG ~~+NOTA_AUG+3P+SG+FEM~~
L (7)		DEPEND+SIMPLE+NAS+bris+V +W2a+PRET+CONJ+3P+SG ~~+NOTA_AUG+2P+PL~~
L (8)		**DEPEND+SIMPLE+NAS+bris+V +W2a+PRET+CONJ+3P+SG +NOTA_AUG+3P+SG+FEM**

(2) Examples of stem and ending encoding for weak verbs (W1 and W2a), with pres. ind. 3sg. abs. / 2pl. conjunct ending -**θ'**. For each verb form cascades of intermediary surface / lower-level stages (**S$_x$**) are given. Rules that do not alter surface forms, e.g. syncope with *marbaid* (a) and *léicid* (b), are not explicitly listed.

(a) W1 *marbaid* ('kills')

L	**SIMPLE+marb+V+W1+PRES+IND+ABS+3P+SG**		
S$_1$	marbāθ'		
S$_2$	marbaθ'		(rewrite stem vowel to *a*)
S$_3$	marbaiθ'		(insert vowel *i* to mark palatalisation)
S$_4$	marbaiθ		(delete palatalisation markers)
S$_5$	(1) **marbaith**	(2) **marbaid**	(rewrite abstract *θ*)

(b) W2a *léicid* ('lets')

L	**SIMPLE+lēc+V+W2a+PRES+IND+ABS+3P+SG**		
S$_1$	léicīθ'		
S$_2$	l'éic'īθ'		(mark palatalisation)
S$_3$	l'éic'iθ'		(rewrite stem vowel to *i* between pal. cons.)
S$_4$	léiciθ		(delete palatalisation markers)
S$_5$	(1) **léicith**	(2) **léicid**	(rewrite abstract *θ*)

(c) W1 *ad·ella* (‘approaches’, ‘visits’)

L	**(DEPEND+)PROTOT+ad+PV1+ell+V+W1+PRES+IND+CONJ+2P+PL**		
S_1	adellāθ'		
S_2	ad'ellāθ'		(mark (pre-syncope) palatalisation)
S_3	ad'[e]llāθ'		(syncope*)
S_4	ad'llāθ'		(syncope*)
S_5	ad'l'l'āθ'		(post-syncope palatalisation)
S_6	ad'l'l'iθ'		(rewrite stem vowel to *i* between pal. consonants)
S_7	aid'l'l'iθ'		(insert vowel *i* to mark palatalisation)
S_8	aidlliθ		(delete pal. markers)
S_9	aidliθ		conson. coalescence (*d-ll* > *dl*)
S_{10}	(1) **aidlith**	(2) **aidlid**	(rewrite abstract *θ*)

* syncope: vowels in all even, but non-final, syllables are marked with square brackets. If there is no rule counteracting syncope, the brackets and vowel are removed in a next step; otherwise, only the brackets are removed. Although it should be possible to encode both the vowel position subject to syncope and conditions that counteract syncope in one step (rule), this two-step approach was found more intuitive and transparent.

References

Bauer, Bernhard. 2014. *The online database of the Old Irish Priscian glosses*. Indogermanistik Wien. URL http://www.univie.ac.at/indogermanistik/priscian/.

Beesley, K. R., and Karttunen, L. (2003). *Finite State Morphology*. Stanford: CSLI Publications/Center for the Study of Language & Information.

Carney, James (1964). *The poems of Blathmac, son of Cú Brettan: together with the Irish Gospel of Thomas and a poem on the Virgin Mary*. London: Irish Texts Society.

Dereza, Oksana (2016). Building a Dictionary-Based Lemmatizer for Old Irish. *Actes de la conférence conjointe JEP-TALN-RECITAL 2016, volume 6: CLTW,* Paris, France, 4 July 2016 (pp. 12-17). URL https://jep-taln2016.limsi.fr/actes/Actes%20JTR-2016/V06-CLTW.pdf.

Dinneen, Patrick S (1927). *Foclóir Gaedhilge agus Béarla: An Irish-English Dictionary*. Revised edition (1904). Baile Átha Cliath: Irish Texts Society. URL https://celt.ucc.ie//Dinneen1sted.html / http://glg.csisdmz.ul.ie/index.php.

Foclóir Stairiúil na Nua-Ghaeilge (Historical Dictionary of Modern Irish, 1600-2000). URL http://www.fng.ie.

Fomin, Maxim and Gregory Toner (2006). Digitizing a Dictionary of Medieval Irish: the eDIL Project. *Literary and Linguistic Computing* 21(1) (pp. 83-90).

Griffith, Aaron, and David Stifter. 2013. *A dictionary of the Old-Irish glosses in the Milan Codex Ambrosianus C 301 inf.* Institut für Sprachwissenschaft, Universität Wien. URL https://www.univie.ac.at/indogermanistik/milan_glosses/.

Lash, Elliott (2014). *The Parsed Old and Middle Irish Corpus (POMIC)*. Version 0.1. URL https://www.dias.ie/celt/celt-publications-2/celt-the-parsed-old-and-middle-irish-corpus-pomic/.

Meid, Wolfgang [ed.] (1974). *Táin bó Fraích.* Mediaeval and Modern Irish Series 22, 2nd ed. (1967), Dublin: Dublin Institute for Advanced Studies.

Lynn, Teresa (2012). Medieval Irish and Computational Linguistics. *Australian Celtic Journal* 10 (pp. 13-27).

McCone, Kim (1997). *The Early Irish Verb.* Maynooth Monographs 1, 2nd ed. (1987). Maynooth: An Sagart.

Ó Crualaoich, Conchubhar (1999). *Some irregular Syncope Patterns in Old Irish*. PhD thesis. National University of Ireland, Maynooth.

Ó Donnaíle, C. *et al*. *In Dúil Bélrai.* URL https://www2.smo.uhi.ac.uk/sengoidelc/duil-belrai/.

Ó Donnaíle, C. *Bunadas*. URL https://www2.smo.uhi.ac.uk/gaidhlig/faclair/bunadas/.

Piotrowski, Michael (2012). *Natural Language Processing for Historical Texts*. Morgan & Claypool. URL https://doi.org/10.2200/S00436ED1V01Y201207HLT017.

Scannell, Kevin (2009). Standardisation of corpus texts for the NEID. URL http://borel.slu.edu/pub/naaclt09.pdf.

Uí Dhonnchadha, E., and Van Genabith, J. (2006). A Part-of-Speech Tagger for Irish using Finite State Morphology and Constraint Grammar Disambiguation, *LREC 2006*, *Genoa*, *May*, 2006 (pp. 2241-2244). URL http://www.lrec-conf.org/proceedings/lrec2006/.

Uí Dhonnchadha, E. *et al.* (2014). Corpas na Gaeilge (1882-1926). Integrating Historical and Modern Irish Texts. *Proceedings of the LREC 2014 Workshop LRT4HDA: Language Resources and Technologies for Processing and Linking Historical Documents and Archives - Deploying Linked Open Data in Cultural Heritage*, Reykjavik, Iceland, May 2014 (pp. 12-18). URL http://www.lrec-conf.org/proceedings/lrec2014/workshops/LREC2014Workshop-LRT4HDA%20Proceedings.pdf