

BBN–ANG–183 Typography

Lecture 8: Diacritics and ligatures

Zoltán Kiss & Péter Szigetvári

Dept of English Linguistics, Eötvös Loránd University

outline

introduction

diacritics one by one

- acute accent

- grave and double acute accent

- breve, caron

- circumflex, macron

- dieresis and umlaut

- dot, ring

- cedilla, comma, ogonek

- tilde, bar

diacritics in Arabic scripts

ligatures

- character ligatures

- stylistic ligatures

- suppression of stylistic ligatures

sample exam questions

diacritics

- ▶ do not occur in isolation, they are typically “carried” by a letter
- ▶ two ways of producing letters with diacritics
 1. the letter with the diacritic is one unique, precomposed character, i.e., one glyph (with a code of its own): e.g., ISO 8859, Unicode, etc.
 2. composition: adding the diacritic to the letter
 - T_EX/L_AT_EX** are truly compositional: any diacritic can be added to any letter: `\’a` yields `á`, `\’q` yields `q́` (`\’@` yields `©!`)
 - HTML** only imitates compositionality (has mnemonic names for characters): `á` yields `á`, `&qacute;` has no meaning
- ▶ when unavailable diacritics can be replaced either by some other diacritic, the addition of some other character, or even simply omitted: e.g., `ó` → `õ` or `ô`; `á` → `'a`, `a'`, `a1`, `aa`, etc.; **François** as **Francois**
- ▶ a site devoted to diacritics: <http://diacritics.typo.cz/>

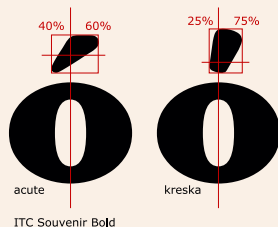
diacritics one by one

acute accent

- ▶ on vowels: marks stress (Italian, Spanish, Portuguese, Welsh), length (Hungarian, Czech, Slovak, Irish), quality (French, Italian, Portuguese), etymology (Polish), high tone (Yoruba), rising tone (Vietnamese), diphthong (Faroese, Icelandic), disambiguation (Dutch, Danish)
- ▶ on consonants: length (Slovak: *ĺ*, *ř*), palatality (Polish, Croatian)



some Polish typographers insist that the acute accent in Polish (a.k.a. kreska) is different from the acute accent — this is probably a source of national pride 😊
(cf. <http://www.twardoch.com/download/polishhowto/kreska.html>)



diacritics one by one

grave (/ɡrɑ:v/) accent

- ▶ on vowels: marks stress (Italian, Catalan), quality (Italian, Catalan, French), shortness (Welsh: *m̀w̃g*), low tone (Yoruba), falling tone (Vietnamese), contracted vowels (Portuguese), disambiguation (French, Italian, Norwegian), pronunciation of the vowel (English: *learnèd*, *blessèd*)



double acute accent

- ▶ is a Hungarian speciality (national pride! 😊)
- ▶ before the 19th c. a combination of acute and umlaut was used: *ő* (cf. <http://www.font.hu/hrant.html>)
- ▶ also used for extra high tone (Tanacross), *ø̊* is used to replace *ø* in Faroese; in linguistic publications a double **grave** accent is used to indicate short low vowels, e.g., *ą̀*, in South Slavic



diacritics one by one

breve

- ▶ on vowels: marks shortness/nonsyllabicity (Cyrillic **ѣ** (= j), cf. **и** (= i)), Esperanto **ŭ** (= w)), different quality (Rumanian **ă** (= ə))
- ▶ on consonants: marks “weakness” (Turkish, Azerbaijani **ğ** (= j))



caron (Czech *háček*, Slovak *mäkčeň*, Serbocroat *kvaka*)

- ▶ on consonants: marks palatality (Czech, Slovak, Serbocroat (note **č** ≠ **ć**))
- ▶ on vowels: marks palatality of the preceding consonant (Czech **ě**), falling rising tone (Mandarin)
- ▶ developed from **č** (along with **š** for length, these are Jan Hus's inventions) — the original shape is still retained in Polish **ż**
- ▶ has a different shape with some consonants: **ř**~**ř**, **ď**~**ď**, **ľ**~**ľ**



Č č Ě ě Š š Ť ť Ď ď Ľ ľ

diacritics one by one

circumflex or caret

- ▶ on vowels: marks length (Welsh *gŵn*), etymological length (Latin *VsC* is often spelt *ŶC* in French: L *costa* ~ F *côte*, OF *beste* ~ F *bête*), diphthong (Slovak *ô* (= [wo])), different quality (Rumanian *â*, earlier *î* (= [u])), openness (Portuguese *ê* (= [ɛ]), *é* (= [e]))
- ▶ on consonants: in Esperanto it is used like the caron: *ĉ* = [tʃ], in Chicheŵa *ŵ* stands for [β]



macron /makrən/

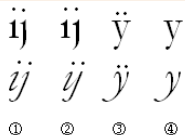
- ▶ on vowels: marks length (Latin, Latvian, Lithuanian), mid tone (Mandarin)
- ▶ on consonants: in medieval scripts it marks doubling: *m̄*, *n̄*, in handwriting it may be used to distinguish *u* and *n*



diacritics one by one

dieresis /dajérəsis/ or trema and umlaut

- ▶ dieresis/trema marks hiatus, i.e., that two vowels are pronounced separately: **naïve**, or that the vowel is pronounced at all: French **aigüe** [ɛgy] (vs. **algu** [alg]), English **Brontë**
- ▶ Albanian **ë** is [ə], Malagasy **ñ** is [ŋ], Dutch **ÿ** is a **ligature** for **ij**
- ▶ umlaut is a different diacritic with the same glyph, marking fronting of a back vowel, in German replacement is possible: **ä/ö/ü** → **ae/oe/ue**
- ▶ heavy metal umlaut: decoration applied for a Gothic feeling, e.g., *Motörhead, Ümlaut*



ij, ij-ligature, ÿ, y schoen, schön, schön

Deutsche Kurrentschrift (1865)

a	<i>a A</i>	j	<i>j J</i>	s	<i>s SP</i>
b	<i>b B</i>	k	<i>k K</i>	t	<i>t T</i>
c	<i>c C</i>	l	<i>l L</i>	u	<i>u U</i>
d	<i>d D</i>	m	<i>m M</i>	v	<i>v V</i>
e	<i>e E</i>	n	<i>n N</i>	w	<i>w W</i>
f	<i>f F</i>	o	<i>o O</i>	x	<i>x X</i>
g	<i>g G</i>	p	<i>p P</i>	y	<i>y Y</i>
h	<i>h H</i>	q	<i>q Q</i>	z	<i>z Z</i>
i	<i>i I</i>	r	<i>r R</i>	tz	<i>tz</i>

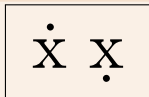
ä Än ö Ön ü Ün
ch ck th sch sz st

= ch, ck, th, sch, sz, st

diacritics one by one

dot

- ▶ Lith. *ė* is long, Maltese *ċ*, *ġ*, *ż* and Pol. *ź* is palatal
- ▶ in transliterations of Semitic scripts *ḏ*, *ṭ*, *ẓ*, *ṣ* represent emphatic (pharyngealized) consonants
- ▶ the **tittle** of lowercase *i* and *j* is taken off when these letters get an upper diacritic: *i* → *ı* → *î* (not *î*)
- ▶ in Turkish *ı* and *İ* are different characters ([*i*] and [*u*]), their uppercase pairs are *İ* and *I*: cf. *İstanbul*, *İzmir*, *Niğde*, *Ağrı*, *İğdir*, *kapı*



ring

- ▶ Danish, Norwegian, Swedish, Walloon *â* is rounded
- ▶ Czech *ů* developed from *o*
- ▶ IPA *ɨ̥*, *ɥ̥* marks voicelessness, in Indo-European studies, syllabicity (note the position of the diacritic depending on the presence of a descender)



diacritics one by one

cedilla

- ▶ developed in Spain from the lower part of a hand-written **z** (z) — hence the name ('little zed') — for “soft” **c**: e.g., **Barcelona** → **Barça**
- ▶ in Turkish, Azerbaijani, etc. **ç** and **ş** occur (for [tʃ ʃ])
- ▶ Rumanian and Latvian have a similar diacritic, a comma: **ș ț** and **ģ ķ ļ ņ**
- ▶ these are often replaced by a cedilla, but careful typographers use a comma

ç

ș

ogonek

- ▶ Polish **ą, ę** and Lithuanian **ą, ę, į, ū** represent historically nasalized vowels; today they are simply long in Lithuanian, variably nasalized in Polish

ą

diacritics one by one

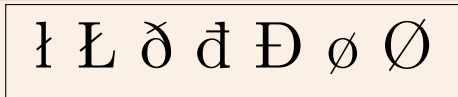
tilde or swung dash

- ▶ developed from an abbreviation mark originally replacing nasals: e.g., **con** → **cõ**, later anything else: e.g., **que** → **q̃**
- ▶ in Portuguese **ãe**, **ão**, and **õe** stand for nasalized diphthongs
- ▶ in Spanish, Galician, Basque **ñ** represents [ɲ]
- ▶ **ḻ** stands for [k̄p̄] in scripts of Vanuatu, in Guaraní **ḡ** is [ŋ]



bar or slash

- ▶ **ł** is [w] in Polish (historically coming from a dark l)
- ▶ in Icelandic and Faroese **ð/Ð** represent [ð] (Scandinavians were close when the IPA was developed)
- ▶ in Serbocroat **đ/Đ** is the voiced counterpart of **ć** ([ɟ̚])
- ▶ in Danish, Faroese, and Norwegian **ø** is [ø] (real close 😊)



the first seven lines of Homer's Iliad

Μῆνιν ἄειδε θεὰ Πηληϊάδεω Ἀχιλλῆος
οὐλομένην, ἣ μυρὶ Ἀχαιοῖς ἄλγε' ἔθηκε,
πολλὰς δ' ἰφθίμους ψυχὰς Ἄϊδι προΐαψεν
ἡρώων, αὐτοὺς δὲ ἐλώρια τεῦχε κύνεσσιν
οἴωνοῖσί τε πᾶσι, Διὸς δ' ἐτελείετο βουλή,
ἔξ οὔ δὴ τὰ πρῶτα διαστήτην ἐρίσαντε
Ἄτρεΐδης τε ἄναξ ἀνδρῶν καὶ δῖος Ἀχιλλεύς.

diacritics in Arabic scripts

short vowels may be rendered by diacritics

بُ بِ بَا بُ بِ بَ ب

individual consonants are often distinguished by diacritics

Near East

Arabic: standard alphabet in which letters of similar shape are brought together:

ب	ت	ث	ج	ح	خ	د	ذ	ر	ز	س	ش	ص
ب	ت	ث	ج	ح	خ	د	ذ	ر	ز	س	ش	ص
ب	ت	ث	ج	ح	خ	د	ذ	ر	ز	س	ش	ص
ب	ت	ث	ج	ح	خ	د	ذ	ر	ز	س	ش	ص

اب پ ت ط ث
 ج چ ح خ د ذ
 ر ز ر ز س ش ص ض
 ظ ع غ ف ق ک گ
 ل م ن و ه ه ع ی

more and more diacritics

قواعد الخط العربي

1. Basic letterforms

قواعد الخط العربي

2. Diacritic Dots

قواعد الخط العربي

3. Vocalization marks in form of vowels

قواعد الخط العربي

4. Decorative elements (without mentioning the numerals, punctuation marks and symbols).

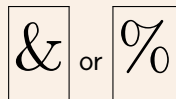


“Allahu akbar” — Allah is great

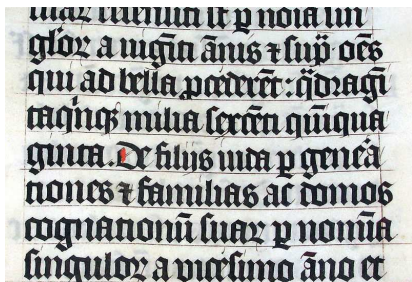
types of ligature

character ligatures

etymologically ligatures, but today considered as a single character: glyphs often occurring together merged into one symbol in scribal conventions (this was a kind of abbreviation), with time they yield a new, “symbiotic” character, e.g.,



some scribal abbreviations



suarum recensiti sunt per nomina sin-
gulorum a viginti annis et supra omnes
qui ad bella procederent: quadragin-
ta quinque milia sescenti quinqu-
ginta. De filiis iuda per genera-
tiones et familias ac domos
cognitionum suarum per nomina
singulorum a vicesimo anno et

Tironian et



survives in

- ▶ Irish
- ▶ the abbreviation *viz.* ← vi+**et** ← videlicet

Tiro was Cicero's scribe; he developed an elaborate system of shorthands, later used by medieval scribes

some character ligatures

- ▶ &: ampersand
- ▶ @: at-sign
- ▶ ¶: pilcrow
- ▶ §: section sign
- ▶ %: percent sign
- ▶ ‰: per mille sign
- ▶ ‰‰: permyriad sign
(basis point)
- ▶ ß: scharfes S, Eszett
- ▶ w/W: double U (or V)
- ▶ æ/Æ, œ/Œ
- ▶ я/Я, ю/Ю: from I + A,
I + O(Y)

& ™ £ ¤ ¥ &

à â á @

ç ¢ € ¶

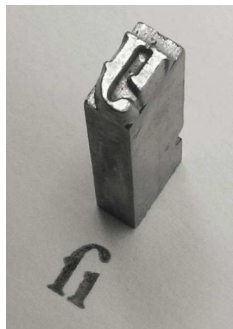
l + s → ß

l + ſ → ſ

l + ð → ð

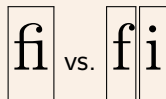


types of ligature



stylistic ligatures

glyphs intruding in each other's bounding box were cast as one glyph (one piece of lead) in the printing press, e.g.,



stylistic ligatures

Five standard latin based ligatures found in most good serif character sets.



"Flow's fine beans" logotype set in un-kerned Mrs Eaves Roman and Italic with no ligatures. Note the collision between the 'fl' of 'flow's' and the 'fi' of 'fine'. Also poor kerning between the apostrophe and letter's either side of it.

The logotype is now kerned and includes ligatures for the 'fl' and the 'fi'. The apostrophe is comfortably set between the w and s and an overall balance is created with the ligatures.

flow's fine beans

flow's fine beans

further stylistic ligatures

fb	fk	ffh	→	fb	fk	ffh		
ff	fl	ffi		ff	fl	ffi		
fh	fr	ffk		fh	fr	ffk	<i>st as ns</i>	
fi	ft	ffl		fi	ft	ffl	<i>ct es nt</i>	
fj	ffb	ffr		fj	ffb	ffr	<i>ty is tt</i>	→
							<i>st as ns</i>	
							<i>ct es nt</i>	
							<i>ty is tt</i>	

The first Manhattan succinct
spelling interpretation contest

æ fb ct fy ee ff gi fh it fj fh it fj ky fl
 gg gy oe sp ggy fr st ft ip py tw tt tw tty
 AA MB © MD ME FF © HE FI UB NK FL LA
 NT © MP E R UD TT UP VA TW UL TY UR
 ffb ffi ffh ffj cky ffl ffr
 fft ffy fi N js fs æ TE Æ

extreme cases of stylistic ligatures

Theremin actress Spokesperson
Krafft risked mistflower octagon
Gottfried outflanked satisfying
Luskin refine bachelor ersatz
Bakerfield jetty factfinding
Nottingham surfboard kaftan
Affects packing offices fjords
Eastfield transforming quota
Vestfjorden Sauerstoffflaschen

TE AARDVARK GØGGLES
Deconstructivist theorists be freeky
 SUPER AMBIENT LAVA LAMP
Affinity with happy gift sprees

Władysław
Jagiello
Władysław
Jagiello

a printer's case with ligatures

A													B																	
A	B	C	D	E	F	G	A	B	C	D	E	F	G	A	B	C	D	E	F	G	A	B	C	D	E	F	G			
H	I	K	L	M	N	O	H	I	K	L	M	N	O	H	I	K	L	M	N	O	H	I	K	L	M	N	O			
P	Q	R	S	T	V	X	P	Q	R	S	T	V	X	P	Q	R	S	T	V	X	P	Q	R	S	T	V	X			
â	è	î	ô	û	Y	Z	J	U	Æ	Æ	ſ	Y	Z	J	U	Æ	Æ	ſ	Y	Z	J	U	Æ	Æ	ſ	Y	Z			
á		í	ó	ú	;	ſb	ffl	ſ	Œ	œ	+	[]	!	ffl	ſ	Œ	œ	+	[]	!	ffl	ſ	Œ	œ	+	[]	!			
à	è	ì	ò	ù	ſt	ſl	ſl	Ç	ç	W	w	()	?	ſl	Ç	ç	W	w	()	?	ſl	Ç	ç	W	w	()	?			
*	ct	J	U	j	ſt	ſl	ff	ë	ï	ü	É	É	”	ff	ë	ï	ü	É	É	”	ff	ë	ï	ü	É	É	”			
a													b																	
o	c	é	-	,		e	1	2	3	4	ſ	6	7	8	1	2	3	4	ſ	6	7	8	1	2	3	4	ſ	6	7	8
&	b	c	d			e	s	f	f	g	h	9	o	s	f	f	g	h	9	o	s	f	f	g	h	9	o			
z						i					ff	ff	k	ſ					ff	ff	k	ſ					ff	ff	k	ſ
y	l	m	n			i	o	p	q		ſ	ſ	:	ſ	o	p	q		ſ	ſ	:	ſ	o	p	q		ſ	ſ	:	ſ
x	v	u	t			Espaces	a	r	.	,			Quadrats	a	r	.	,				Quadrats	a	r	.	,				Quadrats	
D													E																	

avoiding collision

fújhatta magában a csavargók dalát, amly-
gy:

*Teremtette faszekere,
Se feneke, se kereke.
Ne ülj bele, feldűlsz vele.
Sáron-vízen lábad viszen,
Rúgjad a port máris,
Így fut a szamár is.*

stylistic ligatures suppressed

some authorities claim that a stylistic ligature must not span across a word/morpheme boundary, e.g.,

- ▶ Sauerstoffflaschen
- ▶ selfish vs. shellfish

sample exam questions

Which is the cedilla?

1. ç
2. ô
3. č
4. q

Which is not a character ligature?

1. &
2. w
3. @
4. fi