

Devanagari in LuaTeX

Using OpenType features

Ivo Geradts

Kai Eigner

TATZetwerk 

Ondiep Zuidzijde 6, 3551 BW Utrecht

1. About us
2. Foreign scripts in TeX, past and present
3. OpenType features
4. OpenType rendering engines
5. ConTeXt engine
6. Example 1: Latin with mark and mkmk
7. Example 2: Arabic
8. Example 3: Devanagari

1. About us

TAT Zetwerk offers academic typesetting services

Specialties:

- critical editions
- foreign scripts

PlainTeX, for two reasons:

- our exotic needs
- well documented

2.2. Nota etiam contra illam opinionem quam adducit Iohannes¹, scilicet quod scientia de anima est una cum scientia *Parvorum naturalium*, non possunt adduci rationes demonstrative, quamvis ipse adducat nonnullas, que sunt debiles. Sed contra illam positionem potest argui per locum ab auctoritatibus, et maxime Commentatoris primo *Metaphysice* in principio², ubi ponit scientiam *Parvorum naturalium* distinctam esse a scientia libri *De anima*, et per consequens quod aliud est ibi subiectum aliud vero hic.

2.3. Deinde Iohannes³ respondet ad argumenta, et bene.

2.3.1. Et in prima responsione⁴ nota quod illud est de intentione Commentatoris quarti *Metaphysice* in principio⁵, scilicet quod subiectum in scientia non debet esse univocum, sed analogum. Et idem vult sanctus Thomas⁶, licet Scotus⁷ tenuerit oppositum.

2.3.2. Ad tertium respondet Iohannes⁸, sed responsio manifestam patitur instantiam, quia operationes, obiecta et potentie ipsius anime considerantur in isto libro, ut patet; et tamen ille sunt notiores ipsa anima, ut patet secundo *huius*⁹, ubi Philosophus volendo determinare de ipsa anima prius incipit ab obiectis ad operationes, deveniendo postea ad potentias ipsius anime, ultimo ad ipsam animam. Erit ne igitur anima

⁴ contra] contra tenentes animam esse subiectum hic et in libro parvorum naturalium *adn. in marg. V 11* scilicet] subiectum in scientia analogum *adn. in marg. V*

¹ *Locus non inventus.* ² Averroes, *In Metaphysicam IV*, comm. 2, 65^o-66^o. ³ Thomas de Aquino, *Expositio in libros Metaphysicorum*, IV, l. 1, ed. M.-R. Cathala, Torino 1950, 152, n. 544. ⁴ REF. ⁵ Aristoteles, *De anima* II 4, 415a16-21 (AA 6: 56).

¹ Cf. Iohannes de Landino, *Questiones super libros De anima*, l. 3, 13.

² Cf. Iohannes de Landino, *Questiones super libros De anima*, l. 3, 14.

³ Cf. Iohannes de Landino, *Questiones super libros De anima*, l. 3, 14: 'Ad rationes factas in oppositum. Ad primam, cum dicitur "subiectum" etc., dico quod verum est vel unius rationis univoce vel analogice. Et cum dicitur "anima" etc., dico quod verum est univoce, tamen bene est unius analogice, cum per prius dicatur de rationali quam de aliis quare etc.'

⁴ Cf. Iohannes de Landino, *Questiones super libros De anima*, l. 3, 14: 'Ad aliam ("subiectum debet esse" etc.) dico quod verum est inter appropriata scientiæ et non simpliciter, quia tunc ens esset subiectum in omni scientia, cum sit primo notum simpliciter, quod est falsum. Et cum dicitur "anima non est huiusmodi, sed corpus animatum," dico quod corpus animatum non appropriatur isti scientiæ, sed consideratur etiam in scientia parvorum naturalium. Et ideo non valet. Et est simile de hoc sicut de secundis intentionibus consideratis in logica et rebus eis subiectis, quia res sunt notiores nobis quam intentiones, cum notitia intentionum sumatur a rebus, ut dicit Simplicius in *Praedicamentis*. Quia tamen res non appropriantur logicæ, ideo non dicuntur subiectum in logica, sed magis intentiones quare etc.'

notior omnibus istis? Certum est quod non, et maxime loquendo de notitia quoad nos. Et si dices: anima est hic subiectum, et subiectum debet esse notissimum in scientia, ergo etc. – dico quod ista propositio absolute prolata est falsa, scilicet 'subiectum debet esse notissimum in scientia', nec hec invenitur ab aliquo, nec a Philosopho nec Commentatore. Et quod hoc sit verum | demonstro, et maxime in scientiis cuius subiectum sit substantia, quia, cum tale subiectum habeat operationes, proprietates et passiones, nonne ista omnia erunt nobis notiora quam ipsum subiectum? Certe sic, ut patet tertio *huius*¹: non possumus cognoscere substantiam, quia sensus non se profundat usque ad illam; sed bene passiones et operationes eius per sensum possumus percipere; ergo talia erunt nobis notiora. Cum autem anima sit hic subiectum, ut patet, et habeat passiones, operationes etc., que prius sunt nobis note, patet quod illa non erit notior, nec etiam inter appropriata huic scientiæ, quia ut sic ille passiones sunt nobis notiores, ut probatum est. Ideo respondendo ad argumentum duobus modis glosa illam propositionem, scilicet 'subiectum in scientia debet esse primo notum'. Primo: quod si in scientia comparemus subiectum ad passiones et proprietates suas, tunc dico quod subiectum est notissimum in scientia et notior illis notitia propter quid et quoad naturam, non autem quia et quoad nos. Secundo modo glosa illam quod subiectum debet esse primo notum in scientia: verum est de subiecto adequato, comparando illud ad eius partes subiectivas, est notius sicut universaliter magis universale est notius minus universali, ut anima notior est anima sensitiva vel intellectiva etc.

2.3.3. Ad quartum argumentum respondet Iohannes¹. Contra quam responsionem instatur sic: una et eadem passio non potest demonstrari de duobus subiectis realiter distinctis; ideo si aliqua passio demonstrabitur de corpore animato, non poterit demonstrari de anima. Ut capio 'sensibile': vel demonstratur de anima vel de corpore animato. Si de cor-

¹⁶ glosa] quomodo intelligitur propositio subiectum in scientia debet esse primo notum *adn. in marg. V 26* sic] contra Iohannem quod passio non potest de duobus subiectis demonstrari *adn. in marg. V*

¹ REF.

¹ Cf. Iohannes de Landino, *Questiones super libros De anima*, l. 3, 14: 'Ad aliam, cum dicitur "subiectum in scientia" etc., dicitur concedendo etc. Et cum dicitur "anima non est huiusmodi," dico quod, licet passiones anime, ut sunt virtutes et operationes absolute consideratæ, non appropriantur sibi, ut tamen consideratæ sunt in ista scientia sibi appropriantur, quia ista in ista scientia considerantur ut habent ortum ab aliquo principio, et isto modo non appropriantur corpori animato, sed anime etc.'

οὐκ ἄρα μόνον τὴν ἀδελφὴν ἀλλὰ καὶ αὐτὸν ἔδει τυθῆναι, καὶ ἐντεῦθεν ἢ σωτηρία. μετὰ ταῦτα δὲ ἤδη ὑποθέντα τὰ ὀνόματα ἐπεισοδίου ὅπως δὲ ἔσται οἰκεία τὰ ἐπεισόδια, οἷον ἐν τῷ Ὀρέστη ἢ μανία δι' ἧς ἐλήφθη καὶ ἢ σω-
 15 τηρία διὰ τῆς καθάρσεως. ἐν μὲν οὖν τοῖς δράμασι τὰ ἐπεισόδια σύντομα, ἢ δ' ἐποποιία τούτοις μακύνεται. τῆς γὰρ Ὀδυσσεύς οὐ μακρὸς ὁ λόγος ἐστὶν ἀποδημοῦντός τινος ἔτη πολλὰ καὶ παραφυλαττομένου ὑπὸ τοῦ Ποσειδῶνος καὶ μόνου ὄντος, ἔτι δὲ τῶν οἴκοι οὕτως ἐχόντων ὥστε τὰ χρή-
 20 ματα ὑπὸ μνηστήρων ἀναλίσκεσθαι καὶ τὸν υἱὸν ἐπιβουλεύεσθαι, αὐτὸς δὲ ἀφικνεῖται χειμασθεὶς, καὶ ἀναγνωρίσας (εἰς) τινὰς ἐπιθέμενος αὐτὸς μὲν ἐσώθη τοὺς δ' ἐχθροὺς διέφθειρε. τὸ μὲν οὖν ἴδιον τοῦτο, τὰ δ' ἄλλα ἐπεισόδια.
 Ἔστι δὲ πάσης τραγωδίας τὸ μὲν θέσις τὸ δὲ λύσις, τὰ
 25 μὲν ἔξωθεν καὶ ἕνα τῶν ἔσωθεν πολλάκις ἢ θέσις, τὸ δὲ λοιπὸν ἢ λύσις λέγω δὲ θέσιν μὲν εἶναι τὴν ἀπ' ἀρχῆς μέχρι τούτου τοῦ μέρους ὃ ἔσχατόν ἐστιν ἐξ οὗ μεταβαίνει εἰς εὐτυχίαν ἢ εἰς ἀτυχίαν, λύσιν δὲ τὴν ἀπὸ τῆς ἀρχῆς τῆς μεταβάσεως μέχρι τέλους ὥσπερ ἐν τῷ Λυγκεῖ τῷ Θεοδέκτου
 30 θέσις μὲν τὰ τε προπεπραγμένα καὶ ἢ τοῦ παιδίου λήψις καὶ πάλιν ἢ αὐτῶν αἰτίας, λύσις δ' ἢ ἀπὸ τῆς αἰτίας τοῦ θανάτου

11 μόνον Π : μόνην Β : om. Ar. 14 ἐν τῷ AB : in Lat. (τῷ Φ) : τῷ ci. Victorius, Σ ut vid. 15 δράμασι ΒΣ : ἄρμασιν Π 17 οὐ Σ, ci. Vulcanius : om. Ξ ἐστὶν Π : om. Β : [Σ] 19 ἔτι Σ, rec. : ἐπέει Α : in Lat. (= ἐπὶ Φ ci. Minio) : πολλὰ (ex l. 18) Β 20 ὑπὸ μνηστήρων ΦΣ, Paris. 2038 : ὑπομνηστήρων AB 21 δὲ ΠΣ : τὲ (sic) Β : om. Paris. 2038 21-22 ἀναγνωρίσας (εἰς) τινὰς ci. Tarán : ἀναγνωρίσας τινὰς ΠΣ : ἀναγνωρισθεὶς Β 22 ἐπιθέμενος αὐτὸς μὲν Β : αὐτὸς μὲν ἐπιθέμενος Σ : αὐτὸς ἐπιθέμενος αὐτὸς μὲν Π 25 καὶ non vert. Lat. 25-26 ἢ θέσις ... λέγω δὲ om. Σ vel Ar. ex homoioteleuton 27 τούτου non vert. Lat. μεταβαίνει ΦΣ, Paris. 2038 : μεταβαίνειν AB 28 ἢ εἰς ἀτυχίαν ΒΣ : om. Π post λύσιν δὲ Lat. add. esse 29 λυγκεῖ Paris. 2038 : λυκεῖ ΠΣ : γλυκεῖ Β 30 μὲν non vert. Lat. 31 αὐτῶν αἰτίας, λύσις δ' ἢ Σ : αὐτῶν δὴ Ξ αἰτίας Β : αἰτίας Β : [Σ]

μέχρι τοῦ τέλους. τραγωδίας δὲ εἶδη εἰσι τέσσαρα [τοσαῦτα γὰρ καὶ τὰ μέρη ἐλέχθη], ἢ μὲν πεπλεγμένη, ἢς τὸ ὄλον ἐστὶν περιπέτεια καὶ ἀναγνώρισις, ἢ δὲ παθητική, οἷον οἱ τε
 1456a1 Αἰάντες καὶ οἱ Ἴφιονες, ἢ δὲ ἠθική, οἷον αἱ Φθιώτιδες καὶ ὁ Πηλεὺς τὸ δὲ τέταρτον ἢ ἀπλή, οἷον αἱ τε Φορκίδες καὶ ὁ Προ-
 μηθεὺς καὶ ὅσα ἐν ἄδου. μάλιστα μὲν οὖν ἅπαντα δεῖ πει-
 5 ρᾶσθαι ἔχειν, εἰ δὲ μὴ, τὰ μέγιστα καὶ πλείστα, ἄλλως τε καὶ ὡς νῦν συκοφαντοῦσιν τοὺς ποιητὰς γεγονότων γὰρ καθ'
 ἕκαστον μέρος ἀγαθῶν ποιητῶν, ἕκαστου τοῦ ἴδιου ἀγαθοῦ ἀξιοῦσι τὸν ἕνα ὑπερβάλλειν. δίκαιον δὲ καὶ τραγωδίαν ἄλλην καὶ τὴν αὐτὴν λέγειν οὐδενὶ ὡς τῷ μύθῳ τοῦτο
 10 δέ, ὧν ἢ αὐτὴ πλοκὴ καὶ λύσις. πολλοὶ δὲ πλέξαντες εὐλύουσι κακῶς δεῖ δὲ ἀμφοτέρω ἀντικροτεῖσθαι. χρὴ δὲ δι-
 10 περ εἰρηται πολλάκις μεμνήσθαι καὶ μὴ ποιεῖν ἐποποιικὸν σύστημα τραγωδίας—ἐποποιικὸν δὲ λέγω τὸ πολύμυθον—
 οἷον εἰ τις τὸν τῆς Ἰλιάδος ὄλον ποιῶ μύθον. ἐκεῖ μὲν γὰρ διὰ τὸ μήκος λαμβάνει τὰ μέρη τὸ πρέπον μέγεθος, ἐν
 15 δὲ τοῖς δράμασι πολὺ παρὰ τὴν ὑπόληψιν ἀποβαίνει. ση-
 μείον δὲ, ὅσοι πέρσιν Ἰλίου ὄλην ἐποίησαν καὶ μὴ κατὰ μέρος ὥσπερ Εὐριπίδης, (ἢ) Νιόβην καὶ μὴ ὥσπερ Αἰσχύλος, ἢ ἐκπίπτουσιν ἢ κακῶς ἀγωνίζονται, ἐπεὶ καὶ Ἀγάθων ἐξ-

32-33 τοσαῦτα ... ἐλέχθη Ω : secl. Susemihl, ed. pr. 34-1456a1 οἷον ... οἷον] οἷον Σ vel Ar. 2 ἢ ἀπλή ci. Susemihl : ΟΗΣ Β : ὄης Α : spat. vac. in Lat. : non vert. Ar. : ὄης ci. Bywater ὁ Β : om. Α : [Φ] : [Σ] 4 ἄλλως τε Φ, Paris. 2038 : ἄλλως γε Β, fort. Σ : ἀλλ' ὥσπερ (i.e. ὡς γε) Α 6 ἕκαστου Paris. 2038 : ἕκαστον Ω ἴδιου ΑΣ : οἰκείου Β : in proprio Lat. : [Φ] 8 οὐδενὶ ὡς ci. Zeller : οὐδὲν ἴσως ΑΒ : nichil minus Lat. (= οὐδὲν ἴσως?) : οὐδὲν Σ (ἴσως non vert. Ar.) τοῦτο ΠΣ : τούτων Β 10 ἀμφοτέρω ἀντικροτεῖσθαι ci. Immisch : ἀμφοτέρω ἀντικροτεῖσθαι Β : ἀμφοτέρω ἀντικροτεῖσθαι Σ : ἀμφοῦ ἀντικροτεῖσθαι Α : ambo semper coadiunare Lat. (coadiunare ci. Lobel) : ἀμφοῦ ἀντικροτεῖσθαι ci. Vahlen 12 δὲ λέγω ΒΦ, Σ ut vid. : δὲ λέγω δὲ Α 13 ποιοῖ] facit Lat. 15-16 σημείον δὲ ΠΣ : om. Β 16 ὅσοι πέρσιν Ἰλίου ὄλην ΑΣ : ὅσοι πέρσιν Ἰλιούπολιν Β : quicumque quidem Sinilii totam (= ὅσοιπερ συνιλίου ὄλην Φ) 17 ἢ ci. Vahlen : om. Ω 18 Ἀγάθων Α'ΒΦ : ἀγαθῶν ΑΣ

2. Foreign scripts in TeX, past and present

Past: 8-bit PostScript fonts; e.g. approach for classical greek:

- encoding: $\alpha = a$, $\beta = b$, $\gamma = g$
- ligature table: $\sigma = s$, $\zeta = /s$, $\acute{\alpha} = a)$ '
- multitude of font-related files: tfm, vf, pfb, enc

Present: new TeX-engines XeTeX and LuaTeX can handle

OpenType fonts containing Unicode glyph table:

- $\acute{\alpha}$ is unicode position 1F04

3. OpenType features

OpenType fonts contain more information than traditional fonts

- tfm: bounding boxes, kerning, ligatures, etc.
- OpenType additions: GPOS, GSUB, marks

Introduction of XeTeX and LuaTeX removes old limitations:

- vocalized Hebrew, Arabic
- CJK
- Devanagari

4. OpenType rendering engines

OpenType fonts require rendering engine, such as:

- Uniscribe (Windows)
- AAT (OS X); XeTeX on OS X
- Graphite (Windows and Linux); XeTeX on all platforms

- ConTeXt engine: collection of Lua-scripts attached to various callbacks related to font calls and processing of node list

5. ConTeXt engine

- independent of operating system
- readable and modifiable
- work in progress

6. Example 1: Latin script with mark and mkmk

7. Example 2: Arabic

8. Example 3: Devanagari

- Introduction: what is Devanagari?
- Devanagari and OpenType
- Examples

Devanagari

Script

- Sanskrit (old Indic)
- Hindi
- Nepali

Abugida (consonant–vowel sequences are written as a unit)

क = “ka”

क् = “k”

कु = “ku”

क + ◌ (halant)

क + ◌ (matra)

इन्द्रस्य नु वीर्याणि पर वोचं यानि चकार परथमानि वज्री ।
अहन्नहिमन्वपस्ततर्द पर वक्षणा अभिनत पर्वतानाम ॥
अहन्नहिं पर्वते शिश्रियाणं तवष्टास्मै वज्रं सवर्यं ततक्ष ।
वाश्रा इव धेनवः सयन्दमाना अञ्जः समुद्रमव जग्मुरापः ॥
वर्षायमाणो.अव्र्णीत सोमं तरिकद्रुकेष्वपिबत सुतस्य ।
आसायकं मघवादत्त वज्रमहन्नेनं परथमजामहीनाम ॥
यदिन्द्राहन परथमजामहीनामान मायिनाममिनाः परोत मायाः ।
आत सूर्यं जनयन दयामुषासं तादीत्वाशत्रुं न किला विवित्से ॥
अहन वर्त्रं वर्त्रतरं वयंसमिन्द्रो वज्रेण महता वधेन ।



Mahabharata

Syllable structure

- Consonants: e.g. क (“ka”), ष (“ssa”); concatenated by means of halant: in

वक्षणा (“vaksanaah”) ligature-syllable क्ष (“kssa”) produced by क + ◯ + ष

- Matra's: e.g. ◯ा, ि, ◯ी, ◯ु, ◯ृ, ◯ॆ, ◯ॄ, ◯ँ, ◯ौ, ◯े, ◯ै, ◯ॉ, ◯ो, ◯ी, ◯ौ,

◯ॆ, ◯ॄ

- Independent vowels: e.g. अ, आ, इ, ई, उ, ऊ, ऋ, ॠ, ए, ऐ, ओ, औ, ऋ, ॠ

- Vowel modifiers: e.g. ◯ँ, ◯ं, ◯ः and stress marks: e.g. ◯̇, ◯̈, ◯̉, ◯̊

साम्नि

Word with two syllables: “sa” (स) and “mrti” produced by “ma” (म) +

Ligatures and other substitutions

Half-forms (over 1000): क् त → क्त (“kta”) and क् ष → क्ष (“kssa”)

Reph: र (“ra”) + ◯ + क → र्क

क्क	क्ख	क्क	क्ण	क्त	क्त्य	क्त्र	क्त्र्य	क्त्व	क्न	क्न्य	क्म
kka	kkha	kca	kṇa	kta	ktya	ktra	ktrya	ktva	kna	knya	kma
क्य	क्र	क्र्य	क्ल	क्व	क्व्य	क्ष	क्ष्म	क्ष्य	क्ष्व	क्ख्य	क्ख
kya	kra	krya	kla	kva	kvya	kṣa	kṣma	kṣya	kṣva	khya	khra
ग्य	ग्र	ग्र्य	घ्न	घ्न्य	घ्म	घ्य	घ्र	ङ्क	ङ्क	ङ्क्य	ङ्क्य
gya	gra	grya	ghna	ghnya	ghma	ghya	ghra	ṅka	ṅkta	ṅktya	ṅkya
ङ्क्ष	ङ्क्ष्व	ङ्क्ख	ङ्क्ख्य	ङ्ग	ङ्ग्य	ङ्घ	ङ्घ्य	ङ्घ्र	ङ्ङ	ङ्ङ	ङ्ङ
ṅkṣa	ṅkṣva	ṅkha	ṅkhyā	ṅga	ṅgya	ṅgha	ṅghya	ṅghra	ṅṅa	ṅṅa	ṅṅa
ज्य	च्च	च्च	च्च	च्च	च्च	च्च	च्च	च्च	च्च	च्च	च्च
ṅya	cca	ccha	cchra	cṅa	cma	cya	chya	chra	ṅja	ṅjha	ṅjṅa
झ्य	ज्म	ज्य	ज्र	ज्व	ञ	ञ्म	ञ्य	ञ्च	ञ	ञ्य	ट
jṅya	jma	jya	jra	jva	ñca	ñcma	ñcya	ñcha	ñja	ñjya	ṭa

Character reordering

- - Reorder reph (र + ्र) from beginning to end of syllable:

व ("va") + र ("ra") + ्र + त ("ta") + ्र + र ("ra") + ो ("o") →:

वत्रो

- Reorder ्र to beginning of syllable:

OpenType rendering engine

Based on Devanagari section of Microsoft Typography website

- Analyze text
 - recognize syllables
 - find base vowel
 - classify consonants (reph, half-form, pre-base, below-base, post-base)
- Reorder characters using this analysis (e.g. matra's)
- GSUB: basic shaping forms (e.g. features rphf, half, blwf)
- Reorder characters (e.g. reph and matra's)
- GSUB: other features
- GPOS

Examples