http://wwwlot.let.uu.nl/Research/ltrc/eurotyp/index.htm

European Science Foundation

Programme in Language Typology

Eurotyp

Committee on Computation and Standardization

Dik Bakker, Östen Dahl, Martin Haspelmath,
Maria Koptjevskaja-Tamm,
Christian Lehmann, Anna Siewierska

Eurotyp Guidelines

Central Coordination:

Prof. Dr. Ekkehard König

Institut für Englische Philologie (WE 2)

FB Neuere Fremdsprachliche Philologien

Freie Universität, Gosslerstr. 2-4

D - 14195 Berlin

Fondation Européenne de la Science

1, quai Lezay Marnésia

F - 67080 Strasbourg

(Web version by Rob Goedemans)

0. Contents

1. European languages

1.1. Information on languages

1.2. Genetic affiliation

1.3. Abbreviations of languages

1.4. Language index

2. Terminology

2.1. General format for terms and definitions

2.2. Abbreviations of terms

3. Glossing

3.1. Glossing principles

3.2. Abbreviations of grammatical category labels

4. Primary linguistic data

4.1. File structure

*4.2. Tagging: Coding the linguistic structure of a text

5. Questionnaires

5.1. General observations and guidelines

5.2. Elicitation questionnaires

5.3. Analytical questionnnaires

5.4. Survey of questionnaires used in Eurotyp

6. Processing analytical linguistic data

6.1. Coding a linguistic domain

6.2. The representational language

7. Alphabets and character sets

7.1. Transcription tables

7.2. Character coding systems in use

7.3. Coding diacritics and special characters

8. Eurotyp publications style sheet

Style Sheet for Authors and Editors Preparing Conversion Copy

9. Bibliographic entries

9.2. Example entries

* to be filled in

Back to top

1. European languages

1.1. Information on European languages

Based on *Britannica World Data* (1991), *The International Encyclopedia of Linguistics* (1992), *Perepis' SSSR* 1989, and other sources.

A = abbreviation
L = location
S = number of speakers
N = other names
R = remarks
G = references to grammars

Afro-Asiatic

Semitic

1. Assyrian

A: Asr
L: Iraq, Iran, Syria, Turkey, Georgia, Armenia, Azerbaijan
S: 167,000 (reported in 1986)
N: Aisor; Eastern Syriac, Neo-Syrian; Nestorian; Northeastern Aramaic

2. Maltese

A: Mlt
L: Malta
S: more than 400,000 (reported in 1991)
N: Malti

G: Aquilina, Joseph. 1959. *The Structure of Maltese.* Repr. 1973. Malta: Royal University.

French, Edward 1978. *Contemporary journalistic Maltese: an analytical and comparative study.* (Studies in Semitic languages and linguistics VIII) Leiden: E.J. Brill.

Schabert, Peter. 1976. *Laut- und Formenlehre des Maltesischen anhand zweier Mundarten*. Erlangen: Palm & Enke.

Sutcliffe, Edmund F. 1936. *A grammar of the Maltese language with chrestomathy and vocabulary.* London: Oxford University Press.

Altaic

Mongolian

3. Kalmyk

A: Klm
L: Kalmykia (Autonomous Republic within the Russian Federation, on the steppes between the Don and the Volga rivers)
S: 156,400 (reported in 1989)
N: Kalmytz, Kalmuk, Kalmuck, Kalmack, Qalmaq: Oirat-Kalmyk, Kalmyk-Oirat; Western Mongolian

G: Badmaev, B.B. 1966. *Grammatika kalmyckogo jazyka. Morfologija*. Èlista: Kalmyckoe knižnoe izdatel'stvo.

Pjurbeev, G.C. 1977. *Grammatika kalmyckogo jazyka: sintaksis prostogo predloženija.* Èlista: Kalmyckoe knižnoe izdatel'stvo.

Benzing, Johannes. 1985. *Kalmückische Grammatik zum Nachschlagen.* Wiesbaden: Harrassowitz.

Turkic

4. Azerbaijani

A: Azb
L: Azerbaijan (former USSR), West and East Azerbaijan (provinces in Iran), Iraq, Turkey, Syria and Afghanistan
S: 6,614,260 (reported in 1989) in the former USSR, 9,590,000 (reported in 1991) in Iran, and more than 1,000 in the other countries
N: Azeri, Azerbaydzhan

G: Amirpur-Ahrandjani, Manutscher 1971. *Der aserbeidschanische Dialekt von Schahpur. Phonologie und Morphologie.* (Islamkundliche Untersuchungen, Bd.11) Freiburg: Klaus Schwarz.

Fraenkel, Gerd 1962. *A generative grammar of Azerbaijani.* Ph.D. Diss., Indiana University.

Simpson, C.G. 1957. *The Turkish language of Soviet Azerbaijan*. London.

Širaliev, M. Š. & Sevortjan, E.V. (eds.) 1971. *Grammatika azerbajdžanskogo jazyka.* Baku: Izdatel'stvo "Elm".

XXX (ed.) 1971. *Grammatika azerbajdzanskogo jazyka*. Baku.

5. Bashkir

A: Bsh
L: Bashkir Autonomous Republic (Bashkortoston) in the Russian Federation (from the Volga River to beyond the Ural Mountains)
S: 1,047,720 (reported in 1989)
N: Basqort

G: Dmitriev, N.K. 1948. *Grammatika baškirskogo jazyka.* Moskva.

Poppe, Nicholas. 1964. *Bashkir manual*. Bloomington: Indiana University.

Juldasev, A.A. (ed.) 1981. *Grammatika sovremennogo baskirskogo literaturnogo jazyka*. Moskva: Nauka

6. Chuvash

A: Chu
L: Chuvash Autonomous Republic within the Russian Federation (near the Volga River)
S: 1,408,220 (reported in 1989)
N: Bulgar

G: Andreev, N.A. & Egorov, V.G. & Pavlov, I.P. 1957. *Materialy po grammatike sovremennogo Čuvašskogo jazyka. I. Morfologija.* Čeboksary, Chuvash ASSR: Čuvašskij naučno-issledovatel'skij institut jazyka.

Krueger, John Richard. 1961. Chuvash Manual. Bloomington: Indiana University.

7. Crimean Tatar

A: CrTtr
L: formerly spoken in the Crimean Peninsula, from where most speakers were deported to Uzbekistan in 1944; also spoken in Rumania, Bulgaria, Turkey
S: 251,540 (reported in 1989) in the former USSR and more than 25,000 in the other countries
N: Crimean Turkish

8. Gagauz

A: Ggz
L: Moldova, Ukraine, Kazakhstan, Bulgaria and Rumania
S: 173,000 in the former USSR (reported in 1989), and more than 12,000 (reported in 1979) in the other countries
N: Gagauzi

G: Pokrovskaja, L.A. 1964. *Grammatika gagauzskogo jazyka: fonetika i morfologija.* Moskva.

Pokrovskaja, l.A. 1978. *Sintaksis gagauzskogo jazyka v sravnitel'nom otnosenii.* Moskva.

9. Karachai-Balkar

A: Krch
L: Karachay-Cherkes Autonomous Oblast' and Kabardian-Balkar Autonomous Republic (in the Russian Federation)
S: 230,000 (reported in 1989)
N: Karachay-Balkar, Karachai, Karachayla, Karachaylar, Karachayla

G: Aliev, U.B. 1972. *Sintaksis karacaevo-balkarskogo jazyka*. Moskva.

Bajramkulov, U. 1930. *Grammatika karačaevskogo jazyka.* Kislovodsk, Kabardino-Balkar ASSR.

10. Karaim

A: Krm

L: Crimea, southwestern Ukraine, Lithuania
S: 500 (reported in 1989)

G: Prik, O. Ja. 1976. *Očerk grammatiki karaimskogo jazyka. Krymskij dialekt.* Maxačkala: Dagučpedgiz.

Musaev, K.M. 1964. *Grammatika karaimskogo jazyka:fonetika i morfologija.* Moskva.

11. Kumyk

A: Kmk
L: Daghestan, Turkey
S: 274,600 (reported in 1989)
N: Kumuk

G: Dmitriev, N.K. 1940. *Grammatika kumykskogo jazyka.* Moskva-Leningrad.

12. Nogai

A: Nog
L: northern Daghestan and the Cherkes Autonomous Oblast' (in the Russian Federation)
S: 67,600 (reported in 1989)
N: Noghai, Nogay, Noghay

G: Baskakov, N.A. (ed.) 1973. *Grammatika nogajskogo jazyka*. Cerkessk: Karacaevo-cerkesskoe otdelenie stavropol'skogo kniznogo izdatel'stva.

13. Tatar

A: Ttr
L: Tatarston and adjacent areas (within the Russian Federation), Turkey, Afghanistan, China, Finland
S: 5,532,100 (reported in 1989) in the former USSR and more than 7,000 in the other countries

G: Poppe, Nicholas. 1963. *Tatar manual*. Bloomington: Indiana University.

Kurbatov, X.R. et al. (eds) 1969-71. *Sovremennyj tatarskij literaturnyj jazyk.* 2 vols. Moskva.

14. Turkish

A: Trk
L: Turkey, Bulgaria, Greece, Cyprus, Yugoslavia, Rumania, the Ukraine, Russia
S: 50,000,000 (reported in 1987)
N: Turki, Osmanli, Ottoman Turkish

Caucasian

Nakh-Daghestanian

Daghestanian

15. Agul

A: Agl
L: Agul'skij rajon in the southern part of the Daghestanian Republic (in the Russian Federation)
S: 17,700 (reported in 1989)
N: Aghul, Aghulshuy, Aguly

G: Magometov, Aleksandr A. 1970. *Agul'skij jazyk*. Tbilisi: Mecniereba.

16. Akhvakh

A: Axv
L: Axvaxskij rajon in the southern part of the Daghestanian Republic (in the Russian Federation)
S: 5,000 (reported in 1975)
N: Axvax

G: Magomedbekova, Z.M. 1967. *Axvaxskij jazyk*. Tbilisi.

17. Andi

A: And
L: Botlixskij rajon in the southern part of the Daghestanian Republic (in the Russian Federation)
S: 8,000 to 9,000 (reported in 1975)
N: Andii; Qwannab

G: Cercvadze, I. 1965. *Andiuri ena*. Tbilisi.

18. Archi

A: Arc
L: Čarodin'skij rajon in the southern part of the Daghestanian Republic (in the Russian Federation)
S: 859 (reported in 1975)
N: Archin

G: Kibrik, Aleksandr E. et al. 1977. *Opyt strukturnogo opisanija arčinskogo jazyka*. 4 vols. Moskva: Izdatel'stvo MGU.

19. Avar

A: Avr
L: southern part of the Daghestanian Republic (in the Russian Federation) and southern part of Azerbaijan
S: 583,900 (reported in 1989)
N: Dagestani

G: Charachidzé, *Grammaire de la langue avare*. Paris.

Bokarev, Aleksandr A. 1949. *Sintaksis avarskogo jazyka*. Moskva-Leningrad.

20. Bagvalal

A: Bgl
L: Cumadinskij and Axvaxskij rajon in the southern part of the Daghestanian Republic (in the Russian Federation)
S: 5,500 (reported in 1962)
N: Bagulal, Bagval, Bagvalin, Barbalin; Kvanadin, Kvanada

G: Gudava, T. 1967. *Bagvaluri ena*. Tbilisi.

21. Bezhta

A: Bzht
L: Cuntinskij rajon in the southern part of the Daghestanian Republic (in the Russian Federation)
S: 2,500 (reported in 1975), including speakers of Hunzib
N: Bazhita, Bazheta, Bexita, Bechitin; Kapucha, Kupuca, Kapuchin

G: Madieva, G.I. 1965. *Grammatičeskij očerk beztinskogo jazyka*. Maxačkala.

22. Botlikh

A: Btl
L: Botlixskij rajon in the southern part of the Daghestanian Republic (in the Russian Federation)
S: 3,500 (reported in 1962), where the number includes Ghodoberi
N: Botlix

G: Gudava, T. 1963. *Botlixuri ena*. Tbilisi.

23. Budukh

A: Bdx
L: northern Azarbaijan
S: 2,000 (reported in 1977)
N: Budux, Budug, Bukukhi, Budugi

24. Chamalal

A: Chml
L: Cumadinskij rajon in the southern part of the Daghestanian Republic (in the Russian Federation)
S: 5,500 (reported in 1962)
N: Camalal, Chamalin

G: Bokarev, Aleksandr A. 1949. *Očerk grammatiki čamalinskogo jazyka*. Moskva-Leningrad.

25. Dargwa

A: Drgw
L: southern part of the Daghestanian Republic (in the Russian Federation)
S: 355,800 (reported in 1989)
N: Dargin, Dargva; Xjurkilinskij
R: several mutually unintelligible dialects

G: Abdullaev, S.N. 1954. *Grammatika darginskogo jazyka: fonetika i morfologija*. Maxačkala.

Abdullaev, Z.G. 1971. *Očerki po sintaksisu darginskogo jazyka*. Moskva.

Magometov, A.A. 1963. *Kubačinskij jazyk*. Tbilisi.

26. Godoberi

A: Gdb
L: Botlixskij rajon in the southern part of the Daghestanian Republic (in the Russian Federation)
S: ?2,500. Population figures count Ghodoberi together with Botlikh
N: Ghodoberi, Godoberin

G: Saidova, P.A. 1973. *Godoberinskij jazyk.* Maxačkala.

27. Hinukh

A: Hnx
L: Southern part of the Daghestanian Republic (in the Russian Federation)
S: 200 (reported in 1962)
N: Hinux, Ginukh, Ginux

G: Lomtadze, E.A. 1963. *Ginuxskij dialekt didojskogo jazyka*. Tbilisi.

28. Hunzib

A: Hnz
L: Cuntinskij rajon in the southern part of the Daghestanian Republic (in the Russian Federation)
S: ?1000. Population figures count Hunzib together with Bezhta
N: Gunzib; Xunzal, Khunzaly, Khunzal; Enzeb

29. Karata

A: Krt
L: Axvaxskij rajon in the southern part of the Daghestanian Republic (in the Russian Federation)
S: 6,000 (reported in 1962)
N: Karatai; Karain; Kirdi

G: Magomedbekova, Z.M. 1971. *Karatinskij jazyk*. Tbilisi.

30. Khinalug

A: Xnl
L: northern Azerbaijan
S: 1,500
N: Khinalugh, Khinalugi, Xinalug

G: Deseriev, Ju.D. 1959. *Grammatika xinalugskogo jazyka*. Moskva.

Kibrik, Aleksandr E. et al. 1972. *Fragmenty grammatiki xinalugskogo jazyka*. Moskva.

31. Khvarshi

A: Xvr
L: Cumadinskij rajon in the southern part of the Daghestanian Republic (in the Russian Federation)
S: 1,800 (reported in 1962)
N: Khvarshin, Khvarsh, Xvarshi

32. Kryz

A: Krz
L: Northern Azerbaijan
S: 6,000 (reported in 1975)
N: Kryts, Kryzy; Katsy; Dzek, Dzhek, Dzheki

33. Lak

A: Lak
L: southern part of the Daghestanian Republic (in the Russian Federation)
S: 110,470 (reported in 1989)
N: Laki; Kazikumukhtsy

G: Žirkov, L.I. 1955. *Lakskij jazyk: fonetika i morfologija*. Moskva.

34. Lezgian

A: Lzg
L: southern part of the Daghestanian Republic (in the Russian Federation) and northern Azerbaijan
S: 426,640 (reported in 1989)
N: Lezgi, Lezghi, Lezgin; Kiurintsy

G: Uslar, Petr K. 1896. *Etnografija Kavkaza*. VI. Kjurinskij jazyk. Tiflis.

Gadziev, Magomed M. 1954-63. *Sintaksis lezginskogo jazyka*. Vol. 1. 1954. Vol. 2. 1963. Maxačkala.

Haspelmath, Martin. 1993. *A Lezgian grammar*. (Mouton Grammar Library, 9.) Berlin: Mouton de Gruyter.

35. Rutul

A: Rtl
L: Rutul'skij rajon in the southern part of the Daghestanian Republic (in the Russian Federation)
S: 19, 330 (reported in 1989)
N: Rutal; Mykhanidy, Mukhad; Chal

G: Ibragimov, Garun X. 1978. *Rutul'skij jazyk*. Moskva: Nauka.

36. Tabasaran

A: Tbsc
L: Tabasaranskij and Xivskij rajon in the southern part of the Daghestanian

Republic (in the Russian Federation)
S: 93, 550 (reported in 1989)
N: Tabassaran; Ghumghum

G: Magometov, Aleksandr A. 1965. *Tabasaranskij jazyk*. Tbilisi.

Xanmagomedov, 1971. *Očerki po sintaksisu tabasaranskogo jazyka*. Maxačkala.

37. Tindi

A: Tnd
L: Cumadinskij rajon in the southern part of the Daghestanian Republic (in the Russian Federation)
S: 5,000 (reported in 1962)
N: Tindal, Tindin

38. Tsakhur

A: Tsx
L: Rutul'skij rajon in the southern part of the Daghestanian Republic (in the Russian Federation), as well as Zakatal'skij and Caxskij rajon in the northern part of Azerbaijan
S: 19,000 (reported in 1989)
N: Tsaxur, Caxur

G: Ibragimov, Garun X. 1990. *Caxurskij jazyk.* Moskva: Nauka.

39. Tsez

A: Did
L: Cuntinskij rajon in the southern part of the Daghestanian Republic (in the Russian Federation)
S: 7,000 (reported in 1962)
N: Tsuntin; Dido, Didoi (this is the Georgian name of Tsez)

40. Udi

A: Udi
L: northwestern Azerbaijan and eastern Georgia
S: 6, 830 (reported in 1989)
N: Udin, Uti

G: Schulze, Wolfgang. 1982. *Die Sprache der Uden in Nord-Azerbajdzan*. Wiesbaden: Harrassowitz.

Nakh

41. Chechen

A: Che
L: Chechnia-Ingushetia (autonomous republic within the Russian Federation), Kazakhstan, Georgia, Jordan

S: 938, 770 (reported in 1989 for the former USSR)
N: Cecen

G: Jakovlev, Nikolaj F. 1940. *Sintaksis čečenskogo jazyka*. Moskva & Leningrad.

Deseriev, Ju. D. 1960. *Sovremennyj čečenskij literaturnyj jazyk, 1: fonetika*. Groznyj.

42. Ingush

A: Ingu
L: Chechnia-Ingushetia (autonomous republic within the Russian Federation)
S: 230, 290 (reported in 1989)
N: Ingus

G: Mal'sagov, Z.K. 1963. *Grammatika ingusskogo jazyka*. 2nd edn. Groznyj.

43. Tsova-Tush

A: Bts
L: Axmetskij rajon in Georgia
S: 2,500 to 3,000 (reported in 1975)
N: Tush; Bats, Batsbi, Batsi (this is the self-designation of Tsova-Tush speakers; however, they prefer to be called Tsova-Tush by outsiders)

G: Dešeriev, Ju. D. 1953. *Bacbijskij jazyk*. Moskva: Izdatel'stvo AN SSSR.

Abkhaz-Adyghean

44. Abaza

A: Abz
L: Karachay-Cherkes Autonomous Oblast' and the Daghestanian Republic in the western part of the northern Caucasus (in the former USSR), as well as in Turkey
S: 31,400 (reported in 1989) in the former USSR and 5,000 to 30,000 in Turkey
N: Abazin, Tapanta, Ashuwa, Bezshagh (?)

G: Genko, A.N. 1955. *Abazinskij jazyk.* Moskva.

45. Abkhaz

A: Abx
L: Abkhazia (Autonomous Republic in Georgia) and Turkey
S: 98,400 (reported in 1989) in Abkhazia and 4,000 in Turkey
N: Abxaz, Abkhazian

G: Aristava, S.K. et al. 1968. *Grammatika abxazskogo jazyka.* Suxum: Alasara.

Hewitt, George B. 1979. *Abkhaz*. Croom Helm Descriptive Grammars.

Spruit, A. 1986. *Abkhaz studies.* Dissertation, University of Leiden.

46. Adyghe

A: Adg
L: Adyghe Autonomous Oblast' and the Cherkes Autonomous Republic in the Caucasus (in the former USSR), Turkey, Jordan, Syria, Iraq, Israel
S: 118, 200 (reported in 1989) in the former USSR, and 155,000 in other countries
N: Adyge, Adygey, Adigei, Adygei, Adyghian, Adygh; Circassian, Lower Circassian, West Circassian; Kiakh, Kjax.

R: Adyghe and Kabardian are sometimes regarded as varieties of a single language, Circassian or Adygh

G: Jakovlev, Nikolaj F. & Asxamaf, D. 1940. *Grammatika adygejskogo literaturnogo jazyka*. Moskva: Izdatel'stvo AN SSSR

Rogava, G.V. & Keraseva, Z.I. 1966. *Grammatika adygejskogo jazyka.* Majkop.

Smeets, Rieks. *Studies in West Circassian phonology and morphology*. Leiden.

47. Kabardian

A: Kbr
L: Karachay-Cherkes Autonomous Oblast' and the Kabardo-Balkar Autonomous Republic in the western part of the northern Caucasus (in the former USSR), as well as in Saudi Arabia, Turkey and USA.
S: 379,840 for Kabardian and 47,320 for Cherkes (reported in 1989) in the former USSR
N: Kabard; Kabardo-Cherkes; Upper Circassian, East Circassian; Beslenei

G: Bagov, P.M. et al. 1970. *Grammatika kabardino-čerkesskogo literaturnogo jazyka*. C. 1. Fonetika i morfologija. 1970. Moskva: Nauka.

Jakovlev, Nikolaj F. 1948. *Grammatika literaturnogo kabardino-čerkesskogo jazyka.* Moskva: Izdatel'stvo AN SSSR.

Kuipers, Aert H. 1960. *Phoneme and morpheme in Kabardian (Eastern Adyghe).* (Janua Linguarum, series minor, 8) The Hague: Mouton & Co.

Rogava, G.V. et al. 1957. *Grammatika kabardino-čerkesskogo literaturnogo jazyka.* Tbilisi: Institut jazykoznanija AN Gruzinskoj SSR.

XXX (ed.) 1957. *Grammatika kabardino-čerkesskogo literaturnogo jazyka.* Moskva.

48. Ubykh

A: Ubx
L: Formerly spoken in the valleys of the Caucasus east of the Black Sea
S: 1 speaker 82 years old (reported in 1984)
N: Ubyx, Oubykh; Pekhi

Kartvelian

49. Georgian

A: Grg
L: Georgia
S: 3,500,000

G: Aronson, Howard J. 1982. *Georgian - A reading grammar*. Columbus, Ohio: Slavica.

Fähnrich, H. 19??. *Kurze Grammatik der georgischen Sprache.* Leipzig: Verlag Enzyklopädie.

Rudenko, B.T. 1972. *Grammatika gruzinskogo jazyka.* (Janua Linguarum, series anastatica, 7.) The Hague: Mouton & Co.

Tschenkéli, Kita. 1958. *Einführung in die georgische Sprache*. Bd. 1-2. Zürich: Amirani.

Vogt, Hans. 1971. *Grammaire de la langue géorgienne.* Oslo: Universitetsforlaget.

50. Laz

A: Laz
L: southern shore of the Black Sea
S: 50,000
N: Chan

G: Anderson, Ralph D. 1963. *A grammar of Laz.* Ph.D. Diss., University of Texas.

51. Megrelian

A: Mngr
L: Georgia
S: 360,000
N: Mingrelian

52. Svan

A: Svn
L: northwestern Georgia
S: 43,000

Indo-European

Albanic

53. Albanian

A: Alb
L: Albania, Kosovo, southern Italy, Greece, Ukraine
S: 5, 298 000 (reported in 1991)

G: Buchholz, Oda & Fiedler, Wilfried. 1987. *Albanische Grammatik*. Leipzig: Verlag Enzyklopädie.

Camaj, Martin 1969. *Lehrbuch der albanischen Sprache.* Wiesbaden: Harrasowitz.

Gurakuqi, Karl. 1967. *Grammatica albanese dell'uso moderno*. Palermo.

Mann, Stuart E. 1944. *A short Albanian grammar.* London: D. Nutt.

Newmark, Leonard. 1982. *Standard Albanian: a reference grammar for students*. Stanford: Stanford University Press.

Pekmezi, Gjerg 1908. *Grammatik der albanischen Sprache, Laut- und Formenlehre.* Vienna.

Armenian

54. (Modern) Armenian

A: Arm
L: Armenia, eastern Turkey, Middle East
S: 6,000,000

R: two written standards, East Armenian and West Armenian

G: Abeghian, A. 1936. *Neuarmenische Grammatik*. Berlin-Leipzig.

Fairbanks, Gordon H. 1958. *Spoken East Armenian*. New York: American Council of Learned Societies.

Gulian, Kevork H. 1957. *Elementary modern Armenian grammar.* New York, NY: Frederick Ungar.

Kogian, S.L. 1949. *Armenian grammar (West dialect)*. Vienna: Mekhitarist Press.

55. Classical Armenian

A: ClArm
L: Armenia, eastern Asia Minor
N: Grabar

Balto-Slavic

Baltic

56.Latvian

A: Ltv
L: Latvia, Russia, Lithuania, Estonia, Belorussia, the Ukraine
S: around 1,550,000 (reported in 1986)
N: Lettish

G: Endzelin, J. 1922. *Lettische Grammatik*. Riga. (Heidelberg, 1923)

Fennel, Trevor Garth & Gelson, Henry. 1980. *A Grammar of modern Latvian*. Vol. 1-3.

The Hague: Mouton.

Lazdi_a, T.B. 1966. *Latvian.* London: English Universities Press.

57. Lithuanian

A: Lith
L: Lithuania
S: around 3,560,000 (reported in 1989)

G: Ambrazas, V. et al. 1985. *Grammatika litovskogo jazyka*. Vilnius: Mokslas.

Dambriunas, L. & Klimas, A. & Schmalstieg, William R. 1972. *Introduction to modern Lithuanian*. Rev. edn. Brooklyn, N.Y.

Schmalstieg, William R. 1988. *A Lithuanian historical syntax.* Columbus, OH: Slavica.

Senn, Alfred. 1966. *Handbuch der litauischen Sprache*. Vol. 1. Grammatik.

58. Old Prussian

A: OPrs
L: East Prussia, attested in religious texts from the 15th and 16th centuries
S: became extinct in the 17th century

G: Endzelin, J. 1944. *Altpreußische Grammatik*. Riga.

Schmalstieg, William R. 1974. *An Old Prussian grammar*. University Park: Pennsylvania State University.

Slavic

59. Belorussian

A: Bylr
L: Belorussia, Poland
S: 7,116,750 (reported in 1989) in Belorussia and 190,000 (reported in 1991) in Poland
N: Byelorussian; White Russian; White Ruthenian

G: Biryla, M.V. (ed.) 1985-86. *Belaruskaja hramatyka*: u 2 c. 1. fanalohija, arfaepija, marfalohija, slovautvarenne, nacisk. 1985. 2. Sintaksis. 1986. Minsk: Navuka i Texnika.

Lomtev, Timafei P. 1956. *Grammatika belorusskogo jazyka.* Moskva: Učpedgiz.

60. Bulgarian

A: Blg
L: Bulgaria, Moldavia, Rumania, Greece, Turkey
S: 9,000,000 (reported in 1986)

G: Beaulieux, Léon. 1950. *Grammaire de la langue bulgare*. 2nde éd., revue et corr. Paris: Institut d'études slaves.

Maslov, Jurij S. 1956. *Očerk bolgarskoj grammatiki.* Moscow: Isdatel'stvo Literatury na inostrannyx Jazykax.

Scatton, Ernest A. 1984. *A reference grammar of Modern Bulgarian.* Columbus, OH: Slavica.

61. Czech

A: Cz
L: Czechia, Slovakia, the Ukraine, Poland, Austria
S: 11,700,000 (reported in 1986)
N: Bohemian

62. Kashubian

A: Ksh
L: Poland (on the left bank of the lower Vistula River, on the coast west of Gdansk, and southwest from Gdynia)
S: 200,000 (reported in 1977)
N: Cassubian

R: often considered a dialect of Polish

G: Lorentz, Friedrich. 1925. *Geschichte der pomoranischen (kaschubischen) Sprache.* Berlin & Leipzig: de Gruyter.

Perkowski, Jan Louis. 1969. *A Kashubian idiolect in the United States*. Bloomington: Indiana University.

63. Macedonian

A: Mcd
L: Macedonia, Greece, Bulgaria, Albania, Canada
S: 2,000,000 (reported in 1986)

G: Lunt, Horace G. 1952. *A grammar of the Macedonian literary language.* Skopje.

Bojic, Vera & Oschlies, Wolf. 1986. *Lehrbuch der makedonischen Sprache*. 2. Aufl. München: Sagner.

64. Polabian

A: Polb
L: along the river Elbe (Germany)
S: became extinct around 1800

65. Polish

A: Pol
L: Poland, Lithuania, the Ukraine, Belorussian, USA
S: 40,500,000 (reported in 1986)

66. Russian

A: Rus
L: Russia and adjacent areas
S: around 155,000,000 first-language speakers (reported in 1979) and 115,000,000 second-language speakers

67. Serbo-Croatian

A: SCr
L: Serbia, Croatia, Bosnia-Hercegovina, Montenegro, Hungary, Austria, Turkey, USA, Canada, Australia
S: 19,000,000 (reported in 1981)
N: Serbo-Croat (preferred in British English), Croato-Serbian
R: There are two written standard varieties, a western variety written in the Roman alphabet ("Croatian") and an eastern variety written in the Cyrillic alphabet ("Serbian")

68. Slovak

A: Slva
L: Slovakia and adjacent areas, USA, Canada, the Ukraine
S: 5,360,000 (reported in 1985)

69. Slovene

A: Slve
L: Slovenia, Italy, Austria, Hungary
S: 2,220,000 (reported in 1985)
N: Slovenian

G: Lencek, Rado L. 1982. *The structure and history of the Slovene language*. Columbus, Ohio: Slavica Publ.

Svane, Gunnar O. 1958. *Grammatik der slowenischen Schriftsprache.* Copenhagen: Rosenkilde & Bagger.

70. Sorbian, Lower

A: LSrb
L: eastern Germany
S: ?
N: Sorabe; Lower Lusatian, Saxon Lusatian; Dolna Luzica; Windisch, Wendish

71. Sorbian, Upper

A: USrb
L: eastern Germany
S: 74,000 (reported in 1976)
N: Sorabe; Upper Lusatian; Windisch, Wendish

G: Fasske, Helmut. 1981. *Grammatik der obersorbischen Schriftsprache der Gegenwart.* Bautzen: Domowina Verlag.

72. Ukrainian

A: Ukr
L: the Ukraine and adjacent areas, Poland, Czechia, Slovakia, Rumania
S: 45,000,000 (reported in 1986)
N: formerly called Little Russian

G: Luckyj, G. & Rudnyc'kyj, Jaroslav B. 1958. *A modern Ukrainian grammar*. Winnipeg.

Medushevsky, A.P. & Zyatkovska, R. 1963. *Ukrainian grammar.* Kiev: Radjans'ka Škola.

Rusanovskij, V.M. (ed.) 1986. *Ukrainskaja grammatika*. Kiev: Naukova Dumka.

Shevelov, George Y. 1963. *The syntax of Modern Literary Ukrainian: The simple sentence.* (Slavistic Printings and Reprintings). The Hague: Mouton.

73. Old Church Slavonic

A: OCS
L: the written standard is based on the dialect of Thessalonike, but Old Church Slavonic was used as a sacred language throughout the Slavic-speaking world
S: attested in numerous religious texts from the 9th century onwards
N: Old Bulgarian

G: Aitzetmueller, Rudolf. 1978. *Altbulgarische Grammatik also Einführung in die slavische Sprachwissenschaft*. Freiburg i. Br.: Weiher.

Lunt, Horace G. 1955. *Old Church Slavonic grammar*. 's-Gravenhage: Mouton.

Schmalstieg, William R. 1983. *Introduction to Old Church Slavic*. 2nd ed. Columbus, ohio: Slavica Publ.

Vaillant, A. *Le vieux slave.*

Celtic

74. Breton

A: Brt
L: Brittany (France)
S: 570.000 (reported in 1991)
N: Brezhoneg

G: Guillevic, A. 1942. *Grammaire bretonne du dialecte de Vannes.* 4thed. Vannes: Librairie Lafolye & J. de Lamarzelle.

Hardie, D.W.F. 1948. *A handbook of Modern Breton (Armorican).* Cardiff: University of Wales Press.

Hemon, Roparz 1966. *Grammaire bretonne.* 5th ed. La Baule: Al Liamm.

La Gléau, René. 1973. *Syntaxe du Breton moderne 1710-1972.* LaBaule: Éditions La Baule.

McKenna, Malachi. 1988. *A Handbook of modern spoken Breton*. Tübingen: Niemeyer.

Press, Ian. 1986. *A Grammar of modern Breton*. [Mouton Grammar Library] Berlin: Mouton de Gruyter.

Trêpous, Pierre. n.d. (ca. 1970). *Grammaire bretonne.* Rennes: Imp. Simon. Galician

75. Cornish

A: Crn
L: extinct since before 1800 as a first language. Formerly spoken in Cornwall, southwest England.
R: currently being revived for cultural purposes

G: Ellis, P. Berresford. 1974. *The Cornish language and its literature*. London [etc.]: Routledge & Paul.

76. Irish

A: Ir
L: Ireland, Northern Ireland (UK)
S: 170.000 (reported in 1991 for Ireland)
N: (Irish) Gaelic; Erse

G: Bammesberger, Alfred. 1982- *A Handbook of Irish.* Vol. 1-3. Heidelberg: Winter.

Christian Brothers, The. 1962. *New Irish Grammar.* Dublin: Fallons.

McCloskey, Michael James. 1978. *A fragment of a grammar of Modern Irish.* (Texas linguistic forum, 12). Austin, TX: University of Texas.

O'Nolan, Gerald. 1934. *The new era grammar of Modern Irish*. Dublin: Educational Company of Ireland.

77. Manx

A: Mnx
L: extinct as a first language during the 20th century. Formerly spoken on the Isle of Manx, UK.
R: a second language for 200 to 300 people; used for some public functions

G: Broderick, George. 1984. *A Handbook of late spoken Manx.* Vol. 1-2. Tübingen: Niemeyer.

Kneen, J.J. 1931. *A grammar of the Manx language.* London: Oxford University Press (Reprint 1973, Douglas: The Manx Gaelic Society).

78. Scottish Gaelic

A: ScGl
L: Scotland (UK), Canada
S: 80.000 (reported in 1991) in the UK and 5.000 (reported in 1971) in Canada
N: Scots Gaelic, Gaelic

G: Calder, George. 1923. *A Gaelic Grammar.* Glasgow. Repr. 1972. Glasgow: Gairm Publ.

Dorian, Nancy C. 1978. *East Sutherland Gaelic: the dialect of the Brora, Golspie, and Embo fishing communities.* Dublin: Dublin Institute for Advanced Studies.

79. Welsh

A: Wls
L: Wales (UK) and Canada
S: 550.000 (reported in 1991) in the UK and 3.160 (reported in 1971) in Canada
N: Cymraeg, Cymric

G: Jones, John Morris. 1955. *A Welsh Grammar*. Oxford: Clarendon Press.

Jones, Morris & Allan R. Thomas. 1977. *The Welsh language: studies in its syntax and semantics.* Cardiff: University of Wales Press.

Williams, Stephen J. 1980. *A Welsh grammar.* Cardiff: University of Wales Press.

Germanic

80. Danish

A: Dan
L: Denmark, Greenland, northern Germany
S: 5,280,000 (reported in 1980)

81. Dutch

A: Dut
L: the Netherlands, Belgium, Surinam
S: more than 21,000,000 (reported in 1991)
N: Nederlands; Hollands; Flemish, Vlaams, Flamand

82. English

A: Eng
L: British Isles, USA, Canada, Australia, New Zealand, South Africa
S: more than 403,000,000 first language speakers and around 397,000,000 second-language speakers (estimated in 1984)

83. Faroese

A: Far
L: the Faroe Islands, Denmark
S: 47,000 (reported in 1978)

G: Krenn, Ernst. 1940. *Färöische Sprachlehre*. Heidelberg: Winter.

Lockwood, W.B. 1964. *An introduction to Modern Faroese.* Copenhagen: Munksgård.

84. Frisian

A: Frs
L: northern Netherlands, northwestern Germany
S: 751,000 (reported in 1976)
N: Frysk or Fries for Western Frisian
R: subdivided into Eastern, Northern and Western Frisian

G: Anglade, J. 1966. *Petit manuel de frison moderne de l'ouest.* Groningen: Wolters.

Sipma, P. 1913. *Phonology and grammar of Modern West Frisian.* London: Oxford University Press (Publications of the Philological Society).

Tiersma, Pieter M. 1985. *Frisian reference grammar*. Dordrecht: Foris Publications.

85. German

A: Grm
L: Germany, Austria, Switzerland, Hungary, Czechoslovakia, Liechtenstein
S: more than 120,000,000 (reported in 1981)
R: many dialects are not mutually comprehensible. Especially the Low German dialect group may be regarded as a separate language

86. Gothic

A: Goth
L: southern Europe
S: Attested in a bible translation of the 4th century. Continued to be spoken in the Crimea, but is now extinct

G: Braune, W. [16]1961. *Gotische Grammatik.* Neu bearbeitet von E.A. Ebbinghaus. Tübingen: Max Niemeyer Verlag.

Krause, Wolfgang. 1968. *Handbuch des Gotischen*. 3. Aufl. München: Beck.

Wright, J. 1910. *Grammar of the Gothic language.* Oxford: Oxford University Press.

87. Icelandic

A: Ice
L: Iceland
S: 250,000 (reported in 1980)

G: Kress, Bruno. 1982. *Isländische Grammatik*. München: Hueber.

Einarsson, Stefan. 1967. *Icelandic: grammar, texts, glossary.* Baltimore, MD: The John Hopkins Press.

88. Luxembourgeois

A: Lux
L: Luxembourg, Belgium
S: 336,000 or more speakers (reported in 1976)

N: Luxemburgian, Luxembourgish, Letzburgisch, Lëtzeburgesch

89. Norwegian

A: Nor
L: Norway
S: 5,000,000 (reported in 1986)
R: two varieties - Bokmål (Riksmål, Dano-Norwegian) and Nynorsk (Landsmål, New Norse)

90. Swedish

A: Swd
L: Sweden, Finland, USA, Canada
S: 10,000,000 (reported in 1986)

91. Yiddish

A: Yid
L: eastern Poland, Lithuania, Ukraine, Belorussia, Germany, Israel, Canada, USA
S: 2,080,000 (reported in 1986)
N: Judeo-German

G: Birnbaum, Solomon Asher. 1979. *Yiddish*. A survey and a grammar. Manchester University Press.

Katz, Dovid. 1987. *Grammar of the Yiddish language*. London: Duckworth.

Greek

92. Classical Greek

A: ClGrk
L: Greece, eastern Mediterranean, Black Sea
N: Ancient Greek

93. Modern Greek

A: Grk
L: Greece, Cyprus, Italy, Turkey, Albania, Egypt, the Ukraine and adjacent areas
S: around 11,500,000 (reported in 1986)
N: Romaic; Neo-Hellenic

94. Pontic

A: Pon
L: Greece (near Athens) and, probably, Turkey
S: ?
R: sometimes considered a dialect of Modern Greek

95. Tsakonian

A: Tsk
L: eastern coast of the Peloponnesos, Greece
S: 10,000 (reported in 1981)
R: sometimes considered a dialect of Modern Greek

Indo-Iranian

Indic

96. Romany

A: Rmny
L: all over Europe and the Near East
S: more than 2,500,000 (reported in 1986); the exact number of speakers is difficult to estimate
N: Gypsy, Romani
R: several varieties of Romany, some of which are not mutually intelligible

G: Ventcel', Tat'jana V. 1988. *Die Zigeunersprache (nordrussischer Dialekt).* 2. Aufl. Leipzig: Verlag Enzyklopädie. (Translation of: Ventcel', T.V. 1964. *Cyganskij jazyk (severorusskij dialekt).* Moskva.)

Iranian

97. Kirmanji

A: Krmn
L: Turkey, Syria, Iran, Armenia, Georgia, Azerbaijan, Lebanon
S: 7,000,000 to 8,000,000 (estimated in 1987)
N: Kurmanji, Kermanji; Northern Kurdish
R: often considered a dialect of (Northern) Kurdish

G: Bakaev, Čerkes Xudoevič. 1973. *Jazyk Kurdov SSSR.* Moskva: Nauka.

Bedir-Khan, Celadet & Roger Lescot. 1970. *Grammaire kurde (dialecte kurmandji).* Paris: Librairie d'Amérique et d'Orient.

Bedir-Khan, Kamuran Ali. 1953. *Langue kurde.* Paris.

Bedir, Paul. 1926. *Grammaire kurde.* Paris: Librairie Orientale P. Geuthner.

Blau, Jean 1975. *Le kurde de Amadiya et de Djabal Sindjar: Analyse linguistique, textes folkloriques, glossaires.* (Travaux de l'Institut d'Etudes iraniennes de l'Université de la Sorbonne Nouvelle). Paris: Librairie C. Klincksieck.

Fossum, Ludwig O. 1919. *A practical Kurdish grammar.* Minneapolis, MN: Augsburg Publishing House.

Mackenzie, David N. 1961-2. *Kurdish dialect studies* (London Oriental Series, 9 & 10). 2 Vols. London: Oxford University Press.

Soane, Ely B. 1913. *Grammar of the Kurmanji or Kurdish language.* London: Luzac &

Co.

98. Ossetic

A: Oss
L: Ossetia (northern Caucasus, Russian Federation) and Georgia
S: 520,100 (reported in 1989)
N: Ossete

G: Abaev, Vasilij Ivanovic.1964. *A grammatical sketch of Ossetic.* Bloomington: Indiana University.

Isaev, M.I. 1966. *Digorskij dialekt osetinskogo jazyka*. Moskva.

99. Talysh

A: Tls
L: southern Azerbaijan and the adjacent areas in Iran
S: 165,000 to 195,000 (estimated in 1982)
N: Talishi, Talesh

100. Tati

A: Tti
L: Azerbaijan, Daghestan
S: ? 22,040 (reported in 1989)
N: Tat; represented by two main varieties - Jewish Tati (Judoe-Tat, Judeo-Tatic, Hebrew Tat, Jewish Tat, Dzuhuric), and Mussulman Tati (Mussulman Tat, Muslim Tat)
R: the so-called Tat dialects in North-western Iran represent, probably, a different language

G: Grjunberg, A.L. 1963. *Jazyk severoazerbajdzanskix tatov.* Leningrad.

Italic

Romance

101. Aragonese

A: Arag
L: Aragon (Spain)
S: ?
R: sometimes considered a dialect of Spanish

102. Asturian

A: Astu
L: Asturia
S: ?
R: sometimes considered a dialect of Spanish

103. Catalan

A: Ctl
L: northeastern Spain, France, Andorra, Italy, USA
S: 8,840,000 (reported in 1976)
N: Catalonian
R: an official regional language in Spain

G: Badia Margarit, Antonio M. 1962. *Gramatica catalana*. T. 1.2. Madrid: Gredos.

Fabra, Ponmpeu. 1964. *Grammaire catalane*. Paris: Les Belles Lettres.

Gili, Joan 1967. *Introductory Catalan grammar.* Oxford: The Dolphin Book Co.

104. Corsican

A: Cors
L: Corsica (France)
S: ?
R: often considered a dialect of Italian

105. Dalmatian

A: Dlm
L: extinct since the late 19th century; formerly spoken on the coast of Yugoslavia
N: Ragusan

106. Franco-Provençal

A: FPrv
L: southeastern France, northeastern Italy
S: ?
R: Franco-Provençal is a term invented by linguists for a number of transitional dialects that are neither French nor Italian

107. French

A: Fr
L: France, Wallonia, Switzerland, Quebec, Louisiana, Haiti, French Guiana
S: 109,000,000 (reported in 1987)

108. Friulian

A: Frln
L: northeast Italy and adjacent areas of the former Yugoslavia
S: 600,000 (reported in 1986)
N: Friulan, Frioulan, Priulian

109. Galician

A: Glc
L: northwestern Spain (Galicia Province) and Portugal
S: 3,170,000

G: Carballo Calero, Ricardo. 1966. *Gramática elemental del Gallego Común.* 2nd ed. Vigo: Galaxia.

110. Italian

A: It
L: Italy, Ticino
S: 55,000,000
R: many unintelligible dialects, held together by a common written standard based on the Tuscan dialect

111. Ladin

A: Ldn
L: northern Italy (South Tyrol and the Dolomites)
S: 30,000 to 35,000 (reported in 1976)
N: Dolomite, Dolomitic; Ladino
R: distinct from Ladino, or Judeo-Spanish in Israel and Turkey

112. Occitan

A: Prv
L: southeastern France, Italy, Monaco
S: 10,200,000 (reported in 1976)
N: the older name was Provençal
R: Occitan has increasing status as a literary language, but no variety is accepted as standard

G: Bec, Pierre. 1967. *La langue occitane.* (Que sais-je? No. 1059). 2nd ed. Paris: Presses Universitaires de France.

Camproux, Charles 1958. *Étude syntaxique des parlers gévaudanais.* Paris: Presses Universitaires de France.

Compan, André. 1965. *La langue niçoise.* Nice: Éditions Tiranty.

Durand, Bruno. 1941. *Grammaire provençale.* 3rd ed. Aix-en-Provence: Labre.

Kelly, Reine Cardaillac. 1973. *A descriptive analysis of Gascon.* (Janua linguarum, series practica, 138). The Hague: Mouton & Co.

Salvat, Joseph. 1973. *Grammaire occitane: des parlers languedociens.* 3rd ed. Toulouse: Privat.

Teulat, Roger. 1976. *Mémento grammatical de l'occitan referentiel*. Sauvagnas: Cap e cap ed. occitanas.

Xavier de Fourvières, Rodolphe Rieux. 1966. *Grammaire provencale suivie d'un guide de conversation*. Avignon: Aubanel.

113. Portuguese

A: Prt
L: Portugal, Brazil, Angola, Mozambique, Guinea Bissau, East Timor

S: 154,000,000 (reported in 1987)

114. Romansh

A: Rmns
L: Graubünden Canton (Switzerland, on the border with Austria and Italy)
S: 65,000 (reported in 1986)
N: Romantsch, Romanche, Rumantsch, Rhaetian, Rhaeto-Romance (this latter term is sometimes applied to the group consisting of Romansh, Ladin, and Friulian)
R: includes varieties called Engadin and Surselvan. One of the national languages of Switzerland

G: Gartner, Theodor. 1973. *Raetoromanische Grammatik*. Repr. 1973. Walluf bei Wiesbaden: Sändig.

Gregor, Douglas Bartlett. 1982. *Romontsch*. Cambridge: Oleander Press.

115. Rumanian

A: Rum
L: Romania, Moldavia, the former Yugoslavia, Bulgaria, Greece, Albania
S: 23,000,000 (reported in 1986)
N: Daco-Rumanian, Romanian
R: the divergent dialects Istro-Rumanian (Istria), Megleno-Rumanian (northern Greece), and especially Arumanian (southern Balkan) are sometimes considered separate languages. The variety of Moldavia /Moldova had a distinct written form based on the Cyrillic alphabet until 1989 and was considered a separate language in the Soviet Union.

116. Sardinian

A: Srd
L: Sardinia (Italy)
S: more than 1,500,000 (reported in 1977)
R: exist in several varieties - Sardinian Campidanese (South Sardinian), Sardinian Gallurese (Northeastern Sardinian), Sardinian Logudorese (Central Sardinian, Sard, or Sardarese), and Sardinian Sassarese (Northwestern Sardinian). Central Sardinian is the second official language used in Sardinia.

G: Wagner, Max Leopold. 1951. *La lingua sarda*. Storia, spirito e forma. Berna: Francke.

117. Spanish

A: Spn
L: Spain, the Canary Islands, Gibraltar, South America, Mexico and Central America, the Caribbean, USA, the Philippines, Equatorial Guinea, Canada, Australia, France, Morocco
S: around 266,000,000 (reported in 1986)
N: Castilian
R: the standard language is based on the Castilian dialect

118. Latin

A: Ltn
L: originally the Latium area of Italy (around Rome), later the Roman empire
R: Latin was long used as a written language throughout most of Europe and exerted heavy influence on many European languages

other Italic

119. Oscan

A: Osc
L: most of southern Italy until 1st centiry BC
S: was still in use at Pompeii until AD 79

120. Umbrian

A: Umb
L: Iguvium (Gubbio) (tabulae iguvinae, the chief document of Umbrian)
S: attested from 350 to 50 BC

Uralic

Finno-Ugrian

121. Estonian

A: Est
L: Estonia, Latvia, Russian Federation
S: 980,000 (reported in 1989) in the former USSR, and around 100,000 in the other countries

G: Harms, Robert T. 1962. *Estonian grammar*. Bloomington: Indiana University.

Tauli, Valter. 1973. *Standard Estonian grammar*. Vol. 1-2. Uppsala.

122. Finnish

A: Fin
L: Finland, Sweden, Estonia, Norway
S: 5,540,000 (reported in 1987)

123. Hungarian

A: Hng
L: Hungary, Rumania and adjacent areas
S: 14,400,000 (reported in 1986)
N: Magyar

124. Ingrian

A: Ingr
L: to the west of St. Petersburg, and in Sweden
S: 302 (reported in 1989) in the Russian Federation, and from 60 to 80 in Sweden
N: Izhor
R: sometimes regarded as a dialect of Karelian

125. Karelian

A: Krl
L: Karelian Autonomous Republic and the adjacent areas within the Russian Federation, as well as Finland
S: 52, 540 (reported in 1989) in the Russian Federation, and 40,000 (reported in 1979) in Finland
R: distinct from the souteastern dialects of Finnish which are sometimes called 'Karelian'; the Ludic (Ljudikovskij) dialect is occasionally considered a separate language

G: Raun, Alo 1964. *Karelian survey* (Research and Studies in Uralic and Altaic languages, project no. 9). Cleveland: OH: Bell & Howell. 77pp.

126. Komi-Permyak

A: KomP
L: Komi-Permyak National Okrug (within the Russian Federation), west of the central Ural Mountains
S: 106,530 (reported in 1989)
N: Permyak
R: a variety of Komi-Zyryan, but has status as a separate literary language

G: Batalova, R.M. et al. 1962. *Komi-permjackij jazyk*. Kudymkar.

127. Komi-Zyryan

A: Kom
L: Komi Autonomous Republic (within the Russian Federation), near the Arctic Ocean
S: 242,500 (reported in 1989)
N: Komi; Zyryan

G: Austerlitz, Robert 1964. *Permian (Votyak-Zyrien) manual.* (Research and Studies in Uralic and Altaic languages, project no. 64). Cleveland: OH: Bell & Howell.

Lytkin, V.I. (ed.) 1955-64. *Sovremennyj komi jazyk.* 2 vols. Syktyvkar.

128. Livonian

A: Lvn
L: Latvia, the Kurland (Courland) peninsula
S: 99 (reported in 1989)
N: Liv

G: Sjögren, Johann Andreas. 1861. *Livische Grammatik*. St.Petersburg.

Kettunen, L. 1938. Grammatische Einleitung. In: L. Kettunen, *Livisches Wörterbuch*. Helsinki.

129. Mari

A: Mar
L: Mari and Bashkir Autonomous Republics within the Russian Federation, on the left bank of the Volga river
S: 773,800 (reported in 1989)
N: Cheremis
R: two written standards, High Mari and Low Mari

G: Gruzov, L.P. 1960. *Sovremennyj marijskij jazyk: fonetika.* Joškar-Ola.

Timofeeva, V.T. 1961. *Sovremennyj marijskij jazyk: sintaksis složnogo predloženija.* Joskar-Ola.

XXX (ed.) 1961. *Sovremennyj marijskij jazyk: morfologija*. Joškar-Ola.

130. Mordvin

A: Mrd
L: Mordvin Autonomous Republic (within the Russian Federation, western Volga region).
S: 773,820 (reported in 1989)
N: Mordva
R: two written standards, Erzya(-Mordvin), or Mordvin-Erzya, and Moksha(-Mordvin), or Mordvin-Moksha.

G: Koljadenkov, M.N. 1959. *Struktura prostogo predlozenija v mordovskix jazykax.* Saransk.

Paasonen, Heikki 1909. *Mordwinische Chrestomathie mit Glossar und grammatikalischem Abriss.* Helsinki: Finnisch-Ugrische Gesellschaft.

Raun, Alo 1964. *Mordvin manual.* (Research and Studies in Uralic and Altaic languages, project no. 39). Cleveland: OH: Bell & Howell.

Zavodova, R.A. & Koljadenkov, M.N. (eds.) 1964. *Grammatika mordovskix (moksanskogo i erzjanskogo) jazykov*.

131. Udmurt

A: Udm
L: Udmurtia (Autonomous Republic within the Russian Federation) and adjacent areas
S: 520,100 (reported in 1989)
N: Votyak

G: Perevoscikov, P.N. (ed.) 1962. *Grammatika sovremennogo udmurtskogo jazyka*. Izevsk.

Vaxruseva, V.M. et al. (eds) 1974. *Grammatika sovremennogo udmurtskogo jazyka.*

Sintaksis sloznogo predlozenija. Izevsk: Udmurtija.

XXX (ed.) 1970. *Grammatika sovremennogo udmurtskogo jazyka*. Sintaksis prostogo predloženija. Izevsk.

132. Vepsian

A: Vps
L: nortwestern Russia, in the triangle formed by the lakes Ladoga, Onega, and Beloe Ozero
S: 6,350 (reported in 1989)
N: Veps

G: Zajceva, M.I. 1981. *Grammatika vepsskogo jazyka*. Leningrad: Nauka.

133. Votian

A: Vtc
L: nortwestern Russia, between Saint Petersburg and Estonia
S: 28 (reported in 1979)
N: Votic, Vote

G: Ariste, Paul. 1968. *A grammar of the Votic language*. Bloomington: Indiana University.

134. Sami

A: Sam
L: northern Scandinavia, northern Russia
S: 20,000
N: Lapp, Lappish, Saami
R: several Sami languages have to be distinguished, at least Northern Sami, Southern Sami, and Eastern Sami

(perhaps up to 11)

G: Collinder, Björn 1949. *The Lappish dialect of Jukkasjarvi: A morphological survey.* Uppsala: Almqvist & Wiksell.

Kert, G.M. 1971. *Saamskij jazyk (kil'dinskij dialekt)*. Leningrad.

Samoyedic

135. Nenets

A: Nnts
L: across a vast area stretching from the White Sea in European Russia to the delta of the Yenisei river in Asia
S: 26,730 (reported in 1989)
N: Yurak, Yurak Samoyed

G: Décsy, Gyula. 1966. *Yurak Chrestomathy* (Uralic and Altaic Series, Vol. 50).

Bloomington, IN: Indiana University.

Kuprijanova, Z.N. et al. 1957. *Neneckij jazyk.* Leningrad: Gosudarstvennoe Učebno-Pedagogičeskoe Izdatel'stvo Ministerstva Prosveščenija RSFSR.

Basque

136. Basque

A: Bsq
L: Basque country (northeastern Spain and southwestern France (département Pyrénées-Atlantiques))
S: 990,000 (reported in 1991)

G: Arotçarena, Abbé. 1951. *Grammaire basque (dialectes navarrolabourdins).* Tours: Maison Mame.

Gavel, Henri 1929. *Grammaire basque. Tome 1: Phonétique, Partiesdu discours autres que le verbe.* Bayonne: Imprimerie du "Courier".

Gavel, Henri & Georges Lacombe. 1937. *Grammaire basque. Tome 2,Premier fascicule: Le verbe.* Bayonne: Imprimerie de la"Presse".

Houghtan, H.P. 1961. *An introduction to the Basque language.* Leiden: Brill.

Lafitte, P. 1944. *Grammaire Basque: Navarro-labourdin littéraire.* Bayonne.

Saltarelli, Mario. 1988. *Basque.* [Croom Helm Descriptive Grammars] London: Croom Helm.

Etruscan

137. Etruscan

A: Etr
L: attested over a large area of central and northern Italy

G: Stoltenberg, H.L. 1950. *Etruskische Sprachlehre mit vollständigem Wörterbuch.* Leverkusen: Gottschalk.

Back to index

1.2. Genetic affiliation

- **Afro-Asiatic**
  - **Semitic**
    - West
      - Central
        - Aramaic
          - *Assyrian*
        - Arabo-Canaanite
          - *Maltese*
- **Altaic**
  - **Altaic Proper**
    - Turkic
      - Common Turkic
        - Western
          - Bashkir
            - *Bashkir*
          - Kumyk-Karachay
            - *Karachay, Kumyk, Karaim*
          - Tatar
            - *Baraba, Crimean Tatar, Tatar*
        - Southern
          - Crimean Turkish
            - *Crimean Turkish*
          - Gagauz
            - *Gagauz*
          - Azerbaijani
            - *Azerbaijani*
          - Turkish
            - *Turkish*
        - Central
          - *Nogai*
      - Bolgar
        - *Chuvash*
    - Oirat-Kalmyk
      - *Kalmyk*
- **Caucasian**
  - **North**
    - Northeast
      - Dagestan
        - Lezgian
          - Lezgian Proper
            - *Kryts, Tsaxur, Udi, Budux, Lezgi, Tabasaran, Rutul, Agul*
          - Archi
            - *Archi*
          - Xinalug
            - *Xinalug*
        - Lak-Dargwa
          - *Lak, Dargwa*
        - Avaro-Andi-Dido
          - Dido
            - Bezhta-Hunzib
              - *Hunzib, Bezhta*
            - Dido-Hinux
              - *Dido, Hinux*
            - Xvarshi
              - *Xvarshi*
          - Andi
            - *Tindi, Godoberi, Karata, Bagulal, Axvax, Botlix, Andi, Chamalal*
          - Avar
            - *Avar*
      - Nax
        - Chechen-Ingush
          - *Ingush, Chechen*
        - Bats
          - *Bats*
    - Northwest
      - Ubyx
        - *Ubyx*
      - Circassian
        - *Kabardian, Adygh*
      - Abxaz-Abaza
        - *Abaza, Abxaz*
  - **South**
    - Georgian
      - *Georgian*
    - Svan
      - *Svan*
    - Zan
      - *Laz, Mingrelian*
- **Indo-European**
  - **Germanic**

- West
  - Continental
    - East
      - *German, Luxembourgeois, Yiddish*
    - West
      - *Dutch*
    - North Sea
      - *English, Frisian, Fering*
  - North
    - West
      - *Norwegian, Faroese, Icelandic*
    - East
      - *Swedish, Danish*
    - Runic
      - *Runic*
  - East
    - *Vandalic, Burgundian, Gothic*

Italic

- Latino-Faliscan
  - Romance
    - Continental
      - Western
        - Italo-Romance
          - Dalmatian
            - *Dalmatian*
          - Italian
            - *Italian*
        - Gallo-Ibero-Romance
          - Ibero-Romance
            - North
              - Western
              - Central
              - Eastern
            - South
              - *Mozarabic*
          - Gallo-Romance
            - North
              - *Franco-Prov French*
            - South
              - *Provencal*
        - Rheato-Romance
          - *Ladin, Friulian, Romansch*
      - Eastern
        - South
          - *Megleno-Rumanian, Arumanian*
        - North
          - *Istro-Rumanian, Rumanian*
    - Sardinian
      - *Sardinian*
  - Latin
    - *Latin*
  - Faliscan
    - *Faliscan*
- Osco-Umbrian
  - *Oscan, Umbrian, Sabellian*

Balto-Slavic

- Baltic
  - East
    - *Latvian, Lithuanian*
  - West
    - *Old Prussian*
- Slavic
  - West
    - Central
      - *Upper Sorbian, Lower Sorbian*
    - South
      - *Czech, Slovak*
    - North
      - *Kashubian, Polish, Polabian*
  - East
    - North
      - *Byelorussian, Russian*
    - South
      - *Ukrainian*
  - South
    - Old Church Slavonik
      - *Old Church Slavonik*
    - East
      - *Bulgarian, Macedonian*
    - West
      - *Serbo-Croatian, Slovene*

	Greek		*Tsakonian, Classical Greek, Greek*			
	Indo-Iranian					
		Iranian				
			Western			
				Northwest		
					Talysh	*Talysh*
					Kurdish	*Kirmanji*
				Tati	*Tati*	
			West Scythian	*Ossetic*		
		Romany	*Romany*			
	Armenian	*Classical Armenian, Armenian*				
	Albanian	*Albanian*				
	Celtic					
		Insular				
			Goidelic	*Manx, Scottish Gaelic, Irish*		
			Brythonic	*Cornish, Breton, Welsh*		
		Continental	*Gaulish*			
Uralic						
	Finno-Ugrian					
		Finnic				
			North Finnic			
				Baltic Finnic	*Finnish, Vepsian, Votic, Olonets, Ingrian, Livonian, Estonian, Karelian, Ludic*	
				Samic	*Eastern Sami, Northern Sami, Southern Sami*	
			Permic	*Komi, Udmurt*		
			Volgaic	*Mari, Mordvin*		
		Ugric				
			Hungarian			
			Ob-Ugric	*Mansi*		
	North	*Nenets*				
Basque	*Basque*					
Etruscan	*Etruscan*					

Back to index

1.3. Abbreviations of languages

Back to index

Abx	Abkhaz
Abz	Abaza
Adg	Adyghe
Agl	Agul
Alb	Albanian
And	Andi
Arag	Aragonese
Arc	Archi
Arm	Armenian
Asr	Assyrian
Astu	Asturian
Avr	Avar
Axv	Akhvakh
Azb	Azerbaijani
Bdx	Budukh
Bgl	Bagvalal
Blg	Bulgarian
Brt	Breton
Bsh	Bashkir
Bsq	Basque
Btl	Botlikh
Bts	Tsova-Tush
Bylr	Belorussian
Bzht	Bezhta
Che	Chechen
Chml	Chamalal
Chu	Chuvash
ClArm	Classical Armenian
ClGrk	Classical Greek
Cors	Corsican
Crn	Cornish
CrTtr	Crimean Tatar
Ctl	Catalan
Cz	Czech
Dan	Danish

Did	Tsez
Dlm	Dalmatian
Drgw	Dargwa
Dut	Dutch
Eng	English
Est	Estonian
Etr	Etruscan
Far	Faroese
Fin	Finnish
FPrv	Franco-Provençal
Fr	French
Frln	Friulian
Frs	Frisian
Gdb	Godoberi
Ggz	Gagauz
Glc	Galician
Goth	Gothic
Grg	Georgian
Grk	Greek
Grm	German
Hng	Hungarian
Hnx	Hinux
Hnz	Hunzib
Ice	Icelandic
Ingr	Ingrian
Ingu	Ingush
Ir	Irish
It	Italian
Kbr	Kabardian
Klm	Kalmyk
Kmk	Kumyk
KomP	Komi-Permyak
Kom	Komi-Zyryan
Krch	Karachai-Balkar
Krl	Karelian
Krm	Karaim
Krmn	Kirmanji
Krt	Karata
Krz	Kryz
Ksh	Kashubian
Lak	Lak

Lat	Latin
Laz	Laz
Lith	Lithuanian
Liv	Livonian
LSrb	Lower Sorbian
Ltv	Latvian
Lux	Luxembourgeois
Lzg	Lezgian
Mar	Mari
Mcd	Macedonian
Mlt	Maltese
Mngr	Megrelian
Mnx	Manx
Mrd	Mordvin
Nnts	Nenets
Nog	Nogai
Nor	Norwegian
OCS	Old Church Slavonic
OPrs	Old Prussian
Osc	Oscan
Oss	Ossetic
Pol	Polish
Polb	Polabian
Pon	Pontic
Prt	Portuguese
Prv	Occitan
Rmns	Romansh
Rmny	Romany
Rtl	Rutul
Rum	Rumanian
Rus	Russian
ScGl	Scottish Gaelic
SCr	Serbo-Croatian
Slva	Slovak
Slve	Slovene
Spn	Spanish
Srd	Sardinian
Svn	Svan
Swd	Swedish
Tbsc	Tabasaran
Tls	Talysh

Tnd	Tindi
Trk	Turkish
Tsk	Tsakonian
Tsx	Tsakhur
Tti	Tati
Ttr	Tatar
Ubx	Ubykh
Udi	Udi
Udm	Udmurt
Ukr	Ukrainian
Umb	Umbrian
USrb	Upper Sorbian
Vps	Veps
Vtc	Votian
Wls	Welsh
Xnl	Khinalug
Xvr	Khvarshi
Yid	Yiddish

1.4. Language index

Language names as standardized in Eurotyp are put in boldface. Numbers refer to section 1.1.

Abaza 44
Abazin 44
Abazintsy 44
Abkhaz 45
Abkhaz 45
Abkhazian 45
Abxaz 45
Adigei 46
Adyge 46
Adygei 46
Adygey 46
Adygh 46
Adyghe 46
Adyghian 46
Aghul 15
Aghulshuy 15
Agul 15
Aguly 15
Aisor 1
Akhvakh 16
Albanian 53
Ancient Greek 93
Andi 17
Andii 17
Archi 18
Archin 18
Armenian 54
Ashuwa 44
Assyrian 1
Avar 19
Axvax 16
Azerbaijani 4
Azerbaydzhan 4
Azeri 4

Bagulal	20
Bagval	20
Bagvalal	20
Bagvalin	20
Barbalin	20
Bashkir	5
Basqort	5
Basque	136
Bats	43
Batsbi	43
Batsi	43
Bazheta	21
Bazhita	21
Bechitin	21
Belorussian	59
Beslenei	47
Bexita	21
Bezhta	21
Bezshagh (?)	44
Bohemian	61
Bokmål	88
Botlikh	22
Botlix	22
Breton	74
Brezhoneg	74
Budug	23
Budugi	23
Budukh	23
Budux	23
Bukukhi	23
Bulgar	6
Bulgarian	60
Bulgarian, Old	73
Byelorussian	59
Camalal	24
Cassubian	62
Castilian	117
Catalan	103
Catalonian	103
Caxur	38
Cecen	41

Chal	35
Chamalal	24
Chamalin	24
Chan	50
Chechen	41
Cheremis	129
Church Slavonic, Old	73
Chuvash	6
Circassian	46
Classical Armenian	55
Classical Greek	92
Cornish	75
Crimean Tatar	7
Crimean Turkish	7
Croato-Serbian	67
Cymraeg	79
Cymric	79
Czech	61
Daco-Rumanian	117
Dagestani	19
Dalmatian	105
Danish	80
Dano-Norwegian	89
Dargin	25
Dargva	25
Dargwa	25
Dido	39
Didoi	39
Dolna Luzica	70
Dutch	81
Dzek	32
Dzhek	32
Dzheki	32
East Circassian	47
Eastern Syriac	1
Engadin	116
English	82
Enzeb	28
Erse	76
Erzya(-Mordva)	131
Estonian	121

Etruscan 137
Faroese 83
Finnish 122
Flamand 81
Flemish 81
Franco-Provençal 106
French 107
Fries 84
Frioulan 108
Frisian 84
Friulan 108
Friulian 108
Frysk 84
Gaelic 76, 78
Gagauz 8
Gagauzi 8
Galician 109
Georgian 49
German 85
Ghodoberi 26
Ghumghum 36
Ginukh 27
Ginux 27
Godoberi 26
Godoberin 26
Gothic 86
Grabar 55
Greek 93
Gunzib 28
Gypsy 96
Hinukh 27
Hinux 27
Hollands 81
Hungarian 123
Hunzib 28
Icelandic 87
Ingrian 124
Ingus 42
Ingush 42
Irish 76
(Irish) Gaelic 76

Italian 110
Izhor 124
Judeo-German 92
Kabard 47
Kabardian 47
Kabardo-Cherkes 47
Kalmack 3
Kalmuck 3
Kalmuk 3
Kalmyk 3
Kalmyk-Oirat 3
Kalmytz 3
Kapucha 21
Kapuchin 21
Karachai 9
Karachai-Balkar 9
Karachay-Balkar 9
Karaim 10
Karain 29
Karata 29
Karatai 29
Karelian 125
Kashubian 62
Katsy 32
Kazikumukhtsy 33
Kermanji 97
Khinalug 30
Khinalugh 30
Khinalugi 30
Khiurkilinskii 25
Khunzal 28
Khunzaly 28
Khvarsh 31
Khvarshi 31
Khvarshin 31
Kiakh 46
Kirdi 29
Kirmanji 98
Kiurinsty 34
Kjax 46
Komi 127

Komi-Permyak 126
Komi-Zyryan 127
Kryts 32
Kryz 32
Kryzy 32
Kumuk 11
Kumyk 11
Kupuca 21
Kurdish 97
Kurmanji 97
Kvanada 20
Kvanadin 20
Ladin 111
Lak 33
Laki 33
Landsmål 88
Lapp 134
Lappish 134
Latin 118
Latvian 56
Laz 50
Lettish 56
Letzburgisch 88
Lëtzeburgesch 88
Lezghi 34
Lezgi 34
Lezgian 34
Lezgin 34
Lithuanian 57
Little Russian 72
Liv 128
Livonian 128
Lower Lusatian 70
Lower Circassian 46
Lower Sorbian 70
Lud(ic) 125
Lusatian 70, 71
Luxembourgeois 88
Luxembourgish 88
Luxemburgian 88
Macedonian 63

Magyar 123
Maltese 2
Malti 2
Manx 77
Mari 129
Meglenitic 112
Megleno-Rumanian 112
Megrelian 51
Mingrelian 51
Moksha(-Mordva) 130
Mordva 130
Mordvin 130
Mordvin-Erzya 130
Mordvin-Moksha 130
Mozarabic 113
Mukhad 35
Mykhanidy 35
Nederlands 81
Nenets 135
Neo-Hellenic 94
Neo-Syrian 1
Nestorian 1
New Norse 88
Nogai 12
Nogay 12
Noghai 12
Noghay 12
Northeastern Aramaic1
Northern Kurdish 97
Norwegian 89
Nynorsk 89
Occitan 112
Oirat-Kalmyk 3
Old Prussian 58
Old Church Slavonic 73
Oscan 119
Osmanli 14
Ossete 98
Ossetic 98
Ottoman Turkish 14
Oubykh 48

Pekhi 48
Permyak 126
Polish 65
Pontic 94
Portuguese 113
Priulian 108
Provençal 112
Qalmaq 3
Qwannab 17
Ragusan 105
Rhaetian 114
Rhaeto-Romance 114
Riksmål 88
Romaic 94
Romanche 114
Romani 96
Romanian 115
Romansch 114
Romansh 114
Romany 96
Rumanian 115
Rumantsch 114
Russian 66
Rutal 35
Ruthenian 66
Rutul 35
Sami 134
Sard 116
Sardarese 116
Sardinian 116
Saxon Lusatian 70
Scots Gaelic 78
Scottish Gaelic 78
Serbo-Croat 67
Serbo-Croatian 67
Slovak 68
Slovene 69
Slovenian 69
Sorabe 71
Sorabe 70
Spanish 117

Surselvan 114
Svan 52
Swedish 90
Tabasaran 36
Tabassaran 36
Talesh 99
Talishi 99
Talysh 99
Tapanta 44
Tat 100
Tatar 13
Tati 100
Tindal 37
Tindi 37
Tindin 37
Tsakhur 38
Tsakonian 95
Tsaxur 38
Tsez 39
Tsova-Tush 43
Tsuntin 39
Turki 14
Turkish 14
Tush 43
Ubykh 48
Ubyx 48
Udi 40
Udin 40
Udmurt 131
Ukrainian 72
Umbrian 120
Upper Sorbian 71
Upper Circassian 47
Uti 40
Veps 132
Vepsian 132
Vlaams 81
Vote 133
Votian 133
Votic 133
Votyak 131

Welsh 79
Wendish 70
Wendish 71
West Circassian 46
Western Mongolian 3
White Russian 59
White Ruthenian 59
Windisch 70
Windisch 71
Xinalug 30
Xunzal 28
Xvarshi 31
Yiddish 91
Yurak Samoyed 135
Yurak 135
Zyryan 127

Back to index

2. Terminology

2.1. General format for terms and definitions

In a terminological network, a term is defined by a set of relations to other terms. The relations are standardized. At least the following relations hold between linguistic concepts:

1. x *(as an individual) is a (kind of)* y

    e.g.: ablative *is a* case

        is a nominal category
        minal category *is a* morphological category
        causative verb *is a* verb

2. x *(as a class) is a* y

    e.g.: adjective *is a* word class

        nasal *is a* phoneme class
        causative verb *is a* verb class

3. x *is a class of* y

    e.g.: word class *is a class of* word

        phoneme class *is a class of* phoneme
        verb class *is a class of* verb

4. x *is a part/element of* y

    e.g.: word order *is part of* syntax

        syntax *is part of* grammar
        grammar *is part of* language system

5. x *is an aspect/property of* y

    e.g.: arbitrariness *is aspect of* linguistic sign

        nominal category *is aspect of* noun
        adequacy *is aspect of* theory
        endocentricity *is aspect of* syntagm

6. x *is an operator of (the operation/process)* y

    e.g.: case *is operator of* case marking

        comparative *is operator of* comparison

quantifier *is operator of* quantification

7. x *is a result of (the operation/process)* y

e.g.: cleft sentence *is result of* clefting

definite description *is result of* determination
ideophone *is result of* sound symbolism

8. x *is a member of (the relation)* y

e.g.: adjunct *is member of* adjunction

subordinate clause *is member of* subordination
attribute *is member of* attribution
allomorph *is member of* allomorphy

9. x *is a manifestation of (the functional domain)* y

e.g.: possessive affix *is manifestation of* possession

case role *is manifestation of* participation
local relator *is manifestation of* spatial orientation

10. x *expresses* y

e.g.: article *expresses* determination

interrogative sentence *expresses* question
case *expresses* case relation
personal affix *expresses* person

11. x *is an object of (the discipline)* y

e.g.: agglutination *is object of* morphological typology

aphasia *is object of* speech pathology
meaning *is object of* semantics
concept *is object of* logic.

12. x *is a representative of (the discipline)* y

e.g.: Bloomfield *is representative of* American structuralism

Saussure *is representative of* European structuralism
Pike *is representative of* tagmemics

13. x *is cross-related to* y

e.g.: adequacy *is cross-related to* naturalness

adjective *is cross-related to* attribute
adposition *is cross-related to* affix
accusative *is cross-related to* direct object

anaphora *is cross-related to* referential identity

These relations have different logical properties. #1, 3 and 4 are transitive. All of them except #13 are subordinative, i.e. they generate a conceptual hierarchy. #13 will become superfluous once the subordinative relations have been enumerated exhaustively and have been made fully explicit.

Once properties are introduced besides relations, some of the relations are seen to be compound. E.g.:

is an object of the discipline (x, y) iff *is a discipline* (y) and *is an object of* (x, y).

If a distinction between sets and non-sets could be made, then the relation *is part/element of* could be differentiated into 1) *is element/subset of* and 2) *is part of*.

Some of these relations are purely logical, others are peculiar to linguistics. Some connect concepts of like logical status (e.g. #1), others concepts of quite different status (e.g. #12).

A major problem is posed by the systematic polysemy of terms like *syntax*:

a. syntax is part of the object area of linguistics:

→ word order *is a part of* syntax;

b. syntax is a discipline

→ word order *is an object of the discipline* syntax.

2.2. Abbreviations of terms

0	submorphemic unit	morphological unit
1	first person	person
12	first person dual inclusive	dual inclusive
12	first and second person	person
1HML	speaker-humble	humble
1HON	speaker-honorific	honorific
2	second person	person
2HML	addressee-humble	humble
2HON	addressee-honorific	honorific
3	third person	person
3HML	referent-humble	humble
3HON	referent-honorific	honorific

A	transitive subject agreement	subject agreement
A	transitive subject	subject
ABL	ablative	case
ABS	absolutive	case
ABSL	absolute	nominal category
abstr	abstract	nominal category
ABSTR	abstract marker	derivational morpheme
ACAUS	anticausative	detransitivizer
ACC	accusative	case
ACCES	accessory	case
ACNNR	action nominalizer	nominalizer
ACR	actor 1	inflectional category
ACT	active	voice
actr	actor 2	semantic role
ADEL	adelative	case
ADESS	adessive	case
ADIT	aditive	case
Adj	adjective	word, part of speech
Adjl	adjectival	
adjn	adjectivalization	derivation, syntactic process
AdjP	adjective phrase	syntactic category, phrase
ADJR	adjectivalizer	
ADM	admonitive	case
AdNCl	adnominal noun clause	noun clause, attribute
Adp	adposition	relator, particle
AdpP	adposition phrase	syntactic category, phrase
AdRlCl	adverbial relative clause	adverbial clause, relative clause
Adv	adverb	word, part of speech
AdvCl	adverbial clause	subordinate clause, adverbial
Advl	adverbial	modifier
Advn	adverbialization	derivation, syntactic process
AdvP	adverb phrase	syntactic category, phrase

ADVR	adverbializer	
ADVRS	adversative	subordinate clause
AFCT	affect 1	affix
Afct	affect 2	syntactic category
aff	affix	morpheme
AFFMT	affirmative	
AG	agentive	case
Ag	agent	actor 2
AGNR	agent nominalizer	nominalizer
Agr	agreement	syntactic technique
AL	alienable	
ALL	allative	case
ALLOC	allocutive	addressee-honorific
AN	animate	nominal category
ANA	anaphoric	
AND	andative	case
ANT	anterior	relative tense
AOR	aorist	aspect
APASS	antipassive	voice
app	apposition	attribute
APPL	applicative	
APPR	apprehensional	subordinate clause
Art	article	word, part of speech
Asp	aspect	verbal category
ASRT	assertive	mood
ASSOC	associative	case
AT	attributor	linker
ATT	attenuative	aktionsart
attid	attitudinal	adverbial
attrib	attributive	
AUG	augmentative	size
AUX	auxiliary	grammatical formative, part of speech
AVERS	aversive	case

B	bound	
BEN	benefactive 1	case
Ben	benefactive 2	verbal derivation
ben	benefactive 3	semantic role
between	word or clitic between two elements	word, clitic
CARD	cardinal number	nominal category
Card	cardinal	numeral
cat	category	concept
CAUS	causative	verbal category
CCnj	coordinative conjunction	conjunction
CdCl	conditional clause	adverbial clause
CdCnj	conditional conjunction	conjunction
CIRC	circumstantial	subordinate clause
Cl	clause	syntactic construction
cldub	clitic doubling	syntactic process
CLF	classifier	
CLn	noun class n	nominal category
CLT	clitic	word
CLVn	verb class n	verbal category, linguistic class
CMP	completive	aspect
CMPR	comparative	nominal category
CmprCl	comparative clause	subordinate clause
Cnj	conjunction	connective
CNTV	conative	mood
CO	coordinator	
COLL	collective	
colloqu	colloquial	
COM	comitative	case
COMM	common	gender
COMP	complementizer 1	subordinator 1, complementizer 2
Comp	complementizer 2	subordinator 2
CompCl	complement clause	complement 2

Compl	complement 1	sentence component
Con	connective	particle, relator
CONC	concessive	mood
COND	conditional	mood
CONN	connector	connective
CONSEC	consecutive	subordinate clause
CONST	construct	nominal category
CONT	continuous	aktionsart
COP	copula 1	copula 2
Cop	copula 2	verb
Corr	correlative	
CprC	comparative construction	syntactic construction
CRAS	crastinal	tense
CRICl	completive relative clause	relative clause
CsCl	concessive clause	adverbial clause
CsCnj	concessive conjunction	conjunction
ctr	contrast	syntagmatic relation
D1	deictic of 1 person	
D12	deictic of 12 person	
D2	deictic of 2 person	
D3	deictic of 3 person	
DAT	dative	case
DE	dual exclusive	dual, exclusive
Decl	declension	
DECL	declarative	mood
def	definite article	article
DEF	definite	
DEFR	deferential	honorific
Degr	degree	
DEL	delative	case
DEM	demonstrative	
deriv	derivational morpheme	morpheme
DES	desiderative	verbal category

descr	descriptive	
DEST	destinative	case
Det	determiner	syntactic category
DetP	determiner phrase	phrase, syntactic category
DETR	detransitivizer	
DI	dual inclusive	dual, inclusive
dial	used in some dialect but not the standard	
DIM	diminutive	size
DIR	directional 1	case
Dir	directional 2	derivational morpheme
DirAdv	directional adverbial	adverbial
DirC	direct case	case
disc	discontinuity	
dispf	dispreference	
DIST	distal	deictic of 1 person
DO	direct object agreement	object agreement
DObj	direct object	object
Dprn	demonstrative pronoun	pronoun
DR	direct	verbal category
DS	different subject	verbal category
DSTR	distributive	nominal category, verbal category
DU	dual	number
DUB	dubitative	mood
DUR	durative	aktionsart
DWNT	downtoner	particle
DYN	dynamic	aspect
EGR	egressive	aktionsart
ELAT	elative	case
EMPH	emphasizer 1	emphasizer 2
Emph	emphasizer 2	grammatical formative
EMPH	emphatic	
encl	enclitic	

epist	epistemic	
EQT	equative	morphological category
ERG	ergative 1	case
Erg	ergative 2	complement
ESS	essive	case
ev	eventive	
eval	evaluative	
Evid	evidential	modality
EXCL	exclamative	mood
EXIST	exist(ential)	verb
EXP	experiential	aspect
exp	experiencer	semantic role
Expl	expletive element	
extra	extraposition	word order
EXTRV	extraversive	transitive
F	feminine	gender
FACT	factitive	
FAM	familiar	pronominal category
fav	favoured	
FIN	finite	
FNL	final position marker	position marker
Fnl	final	
fo	force	semantic role
FOC	focus marker	emphasizer 2
Foc	focusing	emphasis
FocP	focus position	position
FREQ	frequentative	aktionsart
FRM	formal	mood
frm	found only in formal varieties of the language	formal
FUT	future	tense
G	ditransitive "goal"	semantic role
GEN	genitive	case

Gend	gender	nominal class
GER	gerund	nonfinite verb, verbal noun
giv	given	
GivTop	given topic	topic, given
gnr	generic expression	
GNR	generic	
go	goal	semantic role
HAB	habitual	aspect
HCR	hypocoristic	affect
HEST	hesternal	tense
HML	humble	behavior of address
HODFUT	hodiernal future	future
HODPST	hodiernal past	past
HON	honorific	
HORT	hortative	mood
HUM	human	nominal category
HYP	hypothetical	mood
Idprn	indefinite pronoun	pronoun
Idx	index	formal symbol
ILL	illative	case
IMMFUT	immediate future	future
IMP	imperative	mood
IMPF	imperfect	tense
imposs	impossible	
IMPR	impersonal	personal affix
INAL	inalienable	
INAN	inanimate	nominal category
INCH	inchoative	aktionsart
INCONS	inconsequential	mood
IND	indicative	mood
INDEF	indefinite	
Indef	indefinite article	article
INESS	inessive	case

Inf	infinitive	nonfinite verb, verbal noun
INF	infinitive marker	verbal category
infl	inflection	
INFR	inferential	mood
INGR	ingressive	aktionsart
INJ	injunctive	mood
INST	instrumental	case
INSTNR	instrument nominalizer	nominalizer
instr	instrument	semantic role
Int	intensifier	
INT	interrogative	question particle, morphological category
Interr	interrogative word	word, interrogative
Intj	interjection	particle
INTR	intransitive	verbal category
intro	used typically only in introductory sentences	
INTRV	introversive	detransitivizer
INTS	intensive	aktionsart
INV	inverse	verbal category
inv	inversion	word order, movement rule
INVS	invisible	
IO	indirect object agreement	object agreement
IObj	indirect object	object
IPFV	imperfective	aspect
Iprn	interrogative pronoun	pronoun
IPS	impersonal passive	passive
IpSbj	impersonal subject	subject
IRLS	irrealis	mood
IS	intransitive subject agreement	subject agreement
ISbj	intransitive subject	subject
ITER	iterative	aktionsart, verbal plurality
jrl	journalist usage	language level

JUSS	jussive	mood
Juxt	juxtaposed	
LAT	lative	case
Ld	left dislocation	extraposition, topicalization
lex	lexical	
LIG	ligature	
LNK	linker	
LOC	locative	local case
loc	local rest	semantic role
LOCNR	place nominalizer	nominalizer
LOG	logophoric	pronoun
M	masculine	gender
MAL	malefactive	
MAN	manner 1	case
man	manner 2	semantic role
MANNR	manner nominalizer	nominalizer
marg	marginal	
MCl	main clause	clause type
Md	mood	verbal category
Mdl	modal	
MED	medial	deictic of 2 person
MEDP	mediopassive	voice
MEDT	mediative	case
MEDV	medial verb form	
MHUM	masculine personal	personal
MID	middle	voice
middle	middle of clause; area between Aux V	position
Mod	modifier	sentence component
Modif	modification	linguistic operation
morph	morphological	
MRlCl	modal relative clause	adverbial relative clause, modal
MTV	motivative	case

N	noun	word, part of speech
NAbstr	abstract noun	noun
Nagr	nominal agreement	agreement
Nal	alienable noun	noun
NARR	narrative	tense
narrsequ	used typically in narrative sequences	
NB	not bound	
NBprt	body part noun	noun
Nbr	number	nominal category
NCl	noun clause	complement clause
Nclf	numeral classifier	classifier
NCMP	noncompletive	aspect
Ncol	collective noun	noun
NComm	common noun	noun class
NCount	count noun	noun
NEG	negative	
NFIN	nonfinite	verbal category
NFNL	nonfinal position marker	position marker
Nfoc	nonfocus	
NFUT	nonfuture	tense
NHUM	nonhuman	nominal category
Ninal	inalienable noun	noun
NKin	kinship noun	noun
Nl	nominal	syntactic category
NlCas	cased nominal	nominal
NMass	mass noun	noun
NMHUM	nonmasculine personal	personal
nodiff	no difference	
NOM	nominative	case
NP	noun phrase	syntactic category, phrase
NPL	non-plural	number
NProp	proper noun	noun

NPST	nonpast	tense
NR	nominalizer	
NRFUT	near future	future
NRPST	near past	past
NSG	non-singular	number
NSPEC	nonspecific	
NT	neuter	gender
Ntop	nontopic	
Num	numeral	noun
Num+ORD	ordinal numeral	numeral
NumP	numeral phrase	noun phrase
NVCl	non-verbal clause	clause
NVIS	non-visual evidential	evidential
NVOL	nonvolitional	mood
OBJ	object agreement	verbal agreement
Obj	object	complement
OBL	oblique	oblique case
Obl	oblique case	case
OBLG	obligative	mood
oblig	obligatory	syntagmatic relation
OBV	obviative	person
Obvn	obviation	morphological position
occ	occurrence	linguistic unit
OComp	object of comparison	object
OPT	optative	mood
optl	optional	syntagmatic relation
ORD	ordinal	numeral
out	word or clitic outside of the possessive construction	word, clitic
P	transitive patient	complement
P1	initial position corresponding to GB Comp position	position

PART	participle marker	
Part	participle	nonfinite verb, verbal noun
PASS	passive	voice
Pat	patient	undergoer
PATNR	patient nominalizer	nominalizer
PAU	paucal	number
Pclf	possessive classifier	classifier
PE	plural exclusive	plural, exclusive
PEJ	pejorative	affect
periph	periphrastic	
PERL	perlative	case
Pers	person	verbal category, nominal category, pronominal category
PF	perfect	tense
PFV	perfective	aspect
PI	plural inclusive	plural, inclusive
PL	plural	number
PLUP	pluperfect	tense
PNCT	punctual	aktionsart
PO	primary object agreement	object agreement
PObj	primary object	object
POCRAS	postcrastinal	tense
poetic	used only in poetry	style
POHOD	post-hodiernal	
PoP	postposition phrase	phrase
Posp	postposition	adposition
POSS	possessive	
possb	possible	
possm	possessed	
possr	possessor	
post	postposed	
POST	posterior	relative tense
POSTEL	postelative	case

POSTESS	postessive	case
postV	postverbal	position
POT	potential	mood
Pprn	personal pronoun	pronoun
PREC	precative	mood
PRECLUS	preclusive	case
PRED	predicative	
PredP	predicate phrase	phrase
pref	prefix	affix
prefr	preference	
Prep	preposition	adposition
PrepP	preposition phrase	syntactic category, phrase
prev	preverb	morpheme, relator
preV	preverbal	position
PRHEST	pre-hesternal	tense
Prn	pronoun	noun
PROC	processual	
PROCESS	processive	
procl	proclitic	clitic
PROG	progressive	aspect
PROH	prohibitive	mood
PROL	prolative	case
PROPR	proprietive	possessive
PROSP	prospective	aspect
PROX	proximal	deictic of 1 person
PRS	present	tense
PRTV	partitive	case
PRV	privative	case
PRX	proximate	person
ps	personal	
Psaff	personal affix	affix
psd-cl	pseudo cleft sentence	complex sentence
PsPass	personal passive	passive

PST	past	tense
PstnRlCl	postnominal relative clause	relative clause
Ptl	particle	morpheme, word, part of speech
PURP	purposive	case
purp	purpose	semantic role
Q	question particle	particle
QLNR	quality nominalizer	nominalizer
Qnt	quantifier	word, part of speech
QntP	quantifier phrase	phrase 1
QUOT	quotative	
Rcprn	reciprocal 2	pronoun
Rd	disjuncture	extraposition
RDP	reduplicative	submorphemic unit
rec	recipient	semantic role
RECP	reciprocal 1	voice
RECPST	recent past	past
REFR	referential	determination
REL	relative clause marker	
REL	relative	morpheme
RelCl	relative clause	attribute, clause type
REMFUT	remote future	future
REMPST	remote past	past
REP	repetitive	aktionsart
RES	resultative	aspect, aktionsart
restop	resumed topic	topic
restr	restrictive	
RFL	reflexive 1	voice
Rflprn	reflexive 2	pronoun
RFR	referentive	case
rigid	does not allow variation	
RLS	realis	mood
RPRN	relative pronoun	pronoun

RPRT	reportative	
RVRS	reversive	aktionsart
Sbj	subject	sentence component
SBJ	subject agreement	verbal agreement
SCnj	subordinative conjunction	subordinator 2
sec	second	
sem	semantic	
SENS	sensory evidential	evidential
SEQ	sequential	
set	setting	local adverbial
SG	singular	number
SGT	singulative	nominal category
SIM	simultaneity	taxis
SIM	simultaneous	
Size	size	
SMLF	semelfactive	aktionsart
SO	secondary object agreement	object agreement
SObj	secondary object	object
sou	source	semantic role
spa	speech act	
SPEC	specific	
Spec	specifier	
SR	subordinator 1	subordinator 2
SS	same subject	verbal category
STAT	stative	aktionsart
Std	standard of comparison	
stlg	standard language	register
str	stress	suprasegmentalia
SubCl	subordinate clause	clause, clause type
SUBEL	subelative	case
SUBESS	subessive	case
SUBJ	subjunctive	verbal category

SUBL	sublative	case
Subr	subordinator 2	conjunction
suff	suffix	affix
SUP	superlative	nominal category
SUPEL	superelative	case
SUPESS	superessive	case
SUPL	super-lative	case
synth	synthetic passive	passive
T	ditransitive "theme"	semantic role
temp	temporal	temporal deictic
tend	tendency	
TERM	terminative	aktionsart
Tns	tense	verbal category
TOP	topic	
TopP	topic position	position
TR	transitive	verbal category
Trans	transitivity	fundamental relation
TRNSF	transformative	case
TRNSL	translative	case
TRR	transitivizer	
Tshift	topic shift	topicalization
UGR	undergoer 1	inflectional category
ugr	undergoer 2	semantic role
unacc	unaccusative	
unerg	unergative	
UNSPEC	unspecified	inflectional category
V	verb	word, part of speech
V1	verb initial	position
V2	verb second	position
Vagr	verbal agreement	agreement
val	valence	relationality
var	variation	process
Vce	voice	verbal category

VCl	verbal clause	clause
Vdtr	ditransitive verb	transitive verb
VEN	venitive	directional
Vers	version	voice
Vintr	intransitive verb	verb
VIS	visual	evidential
Vl	verbal	
VN	verbal noun	noun
VOC	vocative	case
VOL	volitional	mood
VP	verb phrase	phrase, syntactic category
VR	verbalizer	
Vrfl	reflexive verb	verb
VS	visible	
Vtr	transitive verb	verb
wght	weight	
wh-quest	question word question	interrogative sentence
Word	word	grammatical unit
yn	yes/no question	interrogative sentence

Back to index

3. Glossing

3.1. Glossing principles

The principles proposed here are based on those set out in:

> Lehmann, Christian (1983). "Directions for interlinear morphemic translations." *Folia Linguistica* 16, 1982:193-224.

3.1.1. Rules

R1. An **interlinear morpheme translation** (Imt) is a translation of a text in a language L1 into a string of elements taken from a language L2 (here, one of the official Eurotyp languages), where, ideally, each morpheme of the L1 text is rendered by a morpheme of L2 or a configuration of symbols representing its meaning, and where the sequence of the units of the translation corresponds to the sequence of the morphemes which they render.

R2. The L1 text is generally given in an orthographic or phonemic representation, thus equalling field #2 or 4, respectively, of section 4.1.4. If L1 has a highly complex morphophonemics, there will be, in addition, a **morphemic representation**, equalling field #6. It will then be this line which the Imt translates.

R3. In non-philological publications, every **foreign language** text under grammatical analysis ought to be accompanied by an Imt. Principled exceptions in Eurotyp are English and French. Further possible exceptions include German, Italian, Spanish, Latin, Russian as well as L2 (=L1) and, possibly, languages genetically closely related to L2.

R4. The **primary aim** of an Imt is to make the grammatical, in particular the morphological, structure of the L1 text transparent.

R5. The **degree of detail** displayed by an Imt depends on the purpose it is meant to serve. The following rules specify the properties of a complete Imt. They do not exclude less detailed Imts where they suffice. Cf. R14 for a possibility of underspecifying morpheme separation in L1.

R6. **Lexical morphemes** of L1 are rendered by lexical morphemes of L2.

R7. L1 **roots** (or stems) are not rendered by inflected (nominative or infinitive) forms of L2, but by roots (or stems).

R8. In contradistinction to an idiomatic translation, an Imt is not context-dependent. **Homonymy** is generally resolved in Imts, **polysemy** is not. A polysemous L1 morpheme is generally consistently rendered by its

nearest context-independent L2 equivalent. Distinct glosses for different uses of the same L1 morpheme are admitted if they correspond to conventional grammatical category labels as described in R9, and there does not exist another morpheme in L1 that takes one of these labels. If an L1 morpheme is rendered simultaneously by two L2 alternatives, these are separated by a slash (/).

R9. A **grammatical formative** of L1 will generally be rendered by a label or configuration of labels taken from the grammatical metalanguage and representing its grammatical meaning (category). Such **grammatical category labels** are put in upper case. Each grammatical category label should represent one and only one grammatical term.

An L2 translation of a grammatical formative may be provided if it is a word in L2 (e.g., adposition, pronoun, negator). Certain L2 words shall not be used in Imts; a list for English is provided in section 3.1.2.

R10. R1 entails that each element in an Imt represents the specific meaning (or grammatical function) of the particular L1 element it renders. It does not represent the **grammatical class** (e.g. the category of a morphological paradigm) of the L1 element.

R11. A **submorphemic unit**, e.g. a euphonic element such as an inserted glide or a Kompositionsfugenelement, may be rendered by 0.

R12. In L1 texts and Imts, the **word boundary symbol** is the blank space ( ), and the principal **morpheme boundary symbol** used is the hyphen (-).

R13. Each unit of the L1 text is rendered by at least one unit in the Imt. If there is a **boundary symbol in the L1 text**, there is a corresponding boundary symbol in the Imt. (In particular: there is a blank space, hyphen, angle bracket, plus sign or equal sign in an Imt if and only if there is an identical symbol in the L1 text corresponding to it.)

R14. If there are separate elements in an Imt — no matter whether they are morphemes, grammatical category labels or words — which do have **distinct morpheme counterparts in the L1 text but the latter are not separated**, such morpheme boundaries shown in the Imt but not in the L1 text are represented by a colon (:). This applies also to **portmanteau morphs**.

R15. If there is an element in an Imt which has **no significans in the L1 text**, it is put between parentheses (()).

R16. If a grammatical meaning is expressed by **internal modification** of a morpheme in the L1 text, the Imt contains first the counterpart of the affected morpheme, then a backslash (\), then the elements representing the meaning of the grammatical process.

R17. Elements representing lexical and/or grammatical **components of a single morpheme** in the L1 text are arranged in a line, and a period/full stop (.) is put between them. The period may be omitted in combinations of person and number.

R18. The hyphen as a general purpose morpheme boundary symbol may be partly substituted, in L1 texts and likewise in lmts, by symbols signalling **specific kinds of morpheme concatenation**. Possible special purpose boundary symbols include the plus sign (+) for compounding and derivation, and the equal sign (=) for cliticization.

R19. **Infixes** and **circumfixes** as well as the elements rendering them in an lmt are set off by angle brackets (<> or ><) as their boundary symbols. In the lmt, the equivalent of the discontinuous morpheme precedes the equivalent of the infix or enclosed sequence, respectively. (Multiple circumfixes are intractable.)

R20. If **constituent structure** is to be displayed, square brackets ([]) can be inserted in the lmt. The use of labeled square brackets is discouraged, except for specific points being made in the context. Cf., however, section 4.1.4, #8.

R21. **Syllabification** is inadmissible in morphemically analyzed texts.

R22. There is no **punctuation** in lmts. Parentheses including optional material in the L1 line are also not repeated in the lmt (cf. R15). The only exception is the orthographic hyphen in the L1 text, which is rendered by two hyphens in the lmt.

R23. For each pair of an L1 text word and the set of elements rendering it, the latter is arranged below the former in such a way that they are left-justified. If such an **arrangement** is impossible, the following minimum requirement must be observed: If there is, in an lmt, an equivalent to an element of an L1 text line, it is contained in the line immediately below that line.

R24. 1. lmts are composed in a smaller type than L1 texts. If this is impossible, at least grammatical category labels should be in small capitals.

2. Grammatical category labels are abbreviated, without a period at the end. Ideally, the abbreviations should not be longer than four letters. Abbreviations for standard terms in the grammatical metalanguage are provided in section 3.2.

3. If such components consist of a specific and a generic category, the generic one is omitted from the lmt (cf. R10).

R25. The **distance** between an L1 text line and the line immediately preceding it is greater than that between it and the lmt line belonging to it.

3.1.2. English morphemes excluded from lmts

An lmt must not take advantage of the (accidental) homonymy of L2 morphemes. No bound grammatical or derivational morphemes appear in lmts (cf. R9). Free grammatical morphemes may be used to render free grammatical morphemes. However, use of those in the second column is discouraged:

Word class	instead of	use
Copulas, auxiliaries	*be* *have* (except to mean `possess, own')	Cop, Pass, Prog ... Pf, Oblg ...
Prepositions	*by* *with* *for* *as* *from* *to* *of*	Ag, Erg ... Inst, Com, Assoc ... Ben, Purp ... Eqt, Ess ... Abl, Del ... Dat, All, Dest, Term, Inf ... Gen
Subordinators	*that* *if*	Comp, Sr (, D3) Int, Cond.Sr
Relativizers	*that* *who* *which*	Rel Rel.Hum.Nom ... Rel.NHum.Nom ...

3.1.3. Symbols

L1	Imt	meaning
x y	x y	word boundary between x and y
x-y	x-y	morpheme boundary between x and y
a<x>b	ab<x>	x is an infix in ab

x>a<y	<xy>a	xy is a circumfix around a
x+y	x+y	x and y form a compound or a derivative stem
x=y	x=y	x and y are joined by clisis
z	x/y	x and y are alternative meanings of ambiguous z
xy	x:y	morpheme boundary between x and y not shown in the L1 text
	(x)	x does not have a significans in the L1 text
z	x\y	y is an internal modification of lexeme x
x	[x]	x is a syntactic constituent
x	$[x]_Y$	x is a syntactic constituent of category Y

3.2. Abbreviations of grammatical category labels

The following list contains the abbreviations of the grammatical category labels that may be used in an lmt. They form a subset of the terms listed in section 2.2. Where there is no abbreviation provided, this means that the term should not be used in an lmt. Terms such as `complementizer, connective, linker, subordinator' are mentioned here under the proviso of R10.

In some cases, a comment indicates that the meaning or necessity of the term remains to be clarified in future versions.

Term	**Abbr**	**Comment**
abessive	(prv) (avers)	recommended terms are `privative' and `aversive'
ablative	abl	`from'
absolute	absl	independent (non-incorporated) form of noun
absolutive	abs	case
accusative	acc	case
action nominalizer	acnnr	

active	act	voice
actor	acr	case or verb agreement or cross-reference position for the actor function in an active system
addressee-honorific	2hon	
addressee-humble	2hml	
adelative	adel	local case
adessive	adess	local case
adhortative	hort	mood
aditive	adit	case; unclear whether necessary
adjectiv(al)izer	adjr	
admonitive	adm	
adverbializer	advr	
adversative	advrs	subordinate clause type: `whereas'
affirmative	affmt	opposite to negative; Welsh *-e*; normally unmarked
agentive	ag	`by'
agent nominalizer	agnr	
alienable	al	possessive attribution morpheme
allative	all	`to'
allocutive	alloc	addressee-honorific in Basque
anaphoric	ana	
andative	and	directional
animate	an	
anterior	ant	relative tense
anticausative	acaus	detransitivization by suppression of agent
antipassive	apass	
aorist	aor	perfective past (as opposed to imperfect)
applicative	appl	verbal derivation. Subtypes may be distinguished by APPL.REC, APPL.INST etc.
apprehensional	appr	subordinate clause type: `lest'
assertive	asrt	modality: subtype of declarative: high degree of commitment
associative	assoc	`the man with the red hat'; distinct from proprietive?

attributor	at	morpheme that links an attribute to the head
augmentative	aug	
auxiliary	aux	only admissible for a morpheme which is neutral to particular verbal categories
aversive	avers	local case
benefactive	ben	case, verbal derivation
cardinal	card	numeral
caritive	(prv)	recommended term is `privative'
causative	caus	valency increase (including transitivization) by addition of agent
circumstantial	circ	subordinate clause type
clamative	(excl)	recommended term is `exclamative'
classifier	clf	
clitic	clt	to be used only if no (functional) categorial distinction applies
collective	coll	
comitative	com	
common	comm	gender (either masc. or fem.); cf. `human' and `animate'
comparative	cmpr	degree
complementizer	comp	kind of subordinator
completive	cmp	if there is only one relevant aspectual distinction, call it perfective (vs. imperfective)
conative	cntv	mood
concessive	conc	
conditional	cond	mood
conjunctive participle	(ger)	recommended term is `gerund'
connector	conn	Hittite *nu*
consecutive	consec	subordinate clause
construct	const	construct state: form of a noun
continuous	cont	
copula	cop	
crastinal	cras	tense referring to tomorrow

customary	(hab)	recommended term is `habitual'
dative	dat	
deagentive	(acaus)	recommended term is `anticausative'
debitive	(oblg)	recommended term is `obligative'
declarative	decl	marker of sentence-type; normally unmarked
deferential	defr	
definite	def	
deictic of 1 person	d1	
deictic of 12 person	d12	
deictic of 2 person	d2	
deictic of 3 person	d3	
delative	del	local case: `down from'
demonstrative	dem	
dependent verb form	(subj)	recommended term is `subjunctive'
desiderative	des	
destinative	dest	local case: `up to'? Cf. also `terminative, purposive'.
detransitivizer	detr	see also `anticausative' and `introversive'
different subject	ds	
diminutive	dim	
direct	dr	as opposed to inverse
direct object	do	verb agreement or cross-reference position
directional	dir	`towards'
distal	dist	far from deictic center
distributive	dstr	nominal or verbal category
ditransitive `goal'	g	
ditransitive `theme'	t	
dual	du	
dual exclusive	de	
dual inclusive	di	
dubitative	dub	
durative	dur	aktionsart

dynamic	dyn	vs. stative
egressive	egr	aktionsart?
elative	elat	local case: `out of'
emphatic	emph	e.g., class of pronoun
equative	eqt	1. case/adposition (`as'); 2. feature/morph of the adjective in equative clauses in Celtic
ergative	erg	case, verb agreement or cross-reference position
essive	ess	case: `as'. See also `transformative'.
exclamative	excl	mood
exclusive		recommended terms are `dual exclusive', `plural exclusive'
exist(ential)	exist	grammatical verb
experiential	exp	aspect; e.g., `have you ever tasted tofu?'
extraversive	extrv	transitivization by addition of patient
factitive	fact	
familiar	fam	pronominal category
feminine	f	gender
finite	fin	
first person dual inclusive	12	if treated as a quasi-singular; otherwise `dual inclusive'
focus	foc	morpheme marking the focus
formal	frm	mood
frequentative	freq	aktionsart: multiple times on several occasions
future	fut	
generic	gnr	determination
genitive	gen	
gerund	ger	verbal adverb or converb; cf. `action nominalizer'
gerundive	(oblg)	recommended term is `obligative'
habitual	hab	
habitual-generic		recommended terms are `habitual', `generic'
habitual-past		recommended terms are `habitual', `past'

hesternal	hest	yesterday's past
hodiernal future	hodfut	
hodiernal past	hodpst	
honorific	hon	
hortative	hort	will usually be `imperative'; use hortative only if morphologically unrelated to imperative
human	hum	
humble	hml	comprises `speaker-humble, addressee-humble, referent-humble'
hypocoristic	hcr	subcategory of affect
hypothetical	hyp	mood
illative	ill	local case: `into'
immediate future	immfut	
immediate past	(recpst)	recommended term is `recent past'
imperative	imp	
imperfect	impf	imperfective past (as opposed to aorist)
imperfective	ipfv	
impersonal	impr	to be used in glosses only if distinct from the specific persons
impersonal passive	ips	passive without promotion of anything to subject
inalienable	inal	possessive attribution morpheme
inanimate	inan	
inceptive	(ingr)	recommended term is 'ingressive'
inchoative	inch	become N/Adj
inclusive		recommended terms are `dual inclusive', `plural inclusive'
inconsequential	incons	subordinate clause
indefinite	indef	
indicative	ind	
indirect object	io	verb agreement or cross-reference position
inessive	iness	local case: `inside'
inferential	infr	mood
infinitive	inf	

ingressive	ingr	aktionsart
injunctive	inj	not generally needed
instructive	(man)	case in Finnish; recommended term is `manner'
instrument nominalizer	instnr	derivational morpheme
instrumental	inst	case
intensive	ints	verbal category, often aktionsart
interrogative	int	question particle or morphological category
intransitive	intr	
intransitive subject	s	verbal agreement or cross-reference position only if opposed to both A and P; use SBJ otherwise
intr. subject and tr. object	(abs)	recommended term is `absolutive'
introversive	intrv	detransitivization by suppression of patient
inverse	inv	as opposed to direct; usually a verbal category
invisible	invs	
irrealis	irls	
iterative	iter	aktionsart: several times on one occasion
jussive	juss	3. ps. imperative or dependent mood in Semitic
lative	lat	local case
ligature	lig	
linker	lnk	morpheme linking subconstiuents of a phrase, typically an NP; properly includes `attributor'
locative	loc	local case
logophoric	log	pronominal or verbal
malefactive	mal	verbal derivation
manner	man	
manner nominalizer	mannr	
masculine	m	gender

masculine personal	mhum	
medial	med	medial distance from deictic center
medial verb form	medv	verb form
mediative	medt	adposition/case (`between, among; by means of'); unclear whether needed
mediopassive	medp	
middle	mid	voice which excludes passive voice
motivative	mtv	case, sometimes called `causal'
mutation		not used outside Celtic
narrative	narr	tense
near future	nrfut	after `immediate future'
necessitative	(oblg)	recommended term is `obligative'
negative	neg	
neuter	nt	gender
nominalizer	nr	see also the more specific ones
nominative	nom	
noncompletive	ncmp	
nonfinite	nfin	
nonfuture	nfut	
nonhuman	nhum	
nonmasculine personal	nmhum	
nonpast	npst	
nonsingular	nsg	
nonspecific	nspec	
nonvolitional	nvol	
non-plural	npl	one or two as opposed to plural
non-singular	nsg	> 1; only if there is a plural for > 2
noun class n	cln	
object	obj	verb agreement or cross-reference position
obligative	oblg	
oblique	obl	case
obviative	obv	
optative	opt	

ordinal	ord	numeral category
participle marker	part	
partitive	prtv	case
passive	pass	
past	pst	
patient nominalizer	patnr	
paucal	pau	
pejorative	pej	subcategory of affect
perfect	pf	tense/aspect
perfective	pfv	aspect
pergressive	(perl)	Unclear. If local case, then recommended term is `perlative'.
perlative	perl	local case
place nominalizer	locnr	
pluperfect	plup	past or perfect of a past
plural	pl	
plural exclusive	pe	
plural inclusive	pi	
pluritive	(pl)	plural of a singulative (e.g. Turkana); recommended term is `plural'
polite	(frm)	recommended term is `formal'
positive	(affm)	recommended term is `affirmative'
possessive	poss	possessive adjective, pronoun and cross-reference; not attributor
postcrastinal	pocras	kind of future
postelative	postel	local case
posterior	post	relative tense
postessive	postess	local case
post-hodiernal	pohod	
potential	pot	
precative	prec	mood for requesting; unclear whether needed
preclusive	preclus	case `apart from'?
predicative	pred	predicative form of nominal base; unclear whether needed

present	prs	
preterite	(pst)	recommended term is `past'
pre-hesternal	prhest	
primary object	po	
privative	prv	case: `without'
processive	process	case?
processual	proc	verbal derivation?
progressive	prog	
prohibitive	proh	
prolative	prol	local case: `along'?
proprietive	propr	`provided with, possessing'
prospective	prosp	`going to'; opposite of perfect
proximal	prox	near the deictic center
proximate	prx	vs. obviative
punctual	pnct	aktionsart morpheme (if it exists)
purposive	purp	case and subordinate clause type
quadral		assumed not to exist
quality nominalizer	qlnr	
quotative	quot	morpheme marking direct speech
realis	rls	verbal category typically used to mark a real state of affairs
recent past	recpst	
reciprocal	recp	pronominal or verbal; if the latter, a voice
reduplicative	(rdp)	gloss by function or treat as phonological part of allomorph
referentive	rfr	case: `about'
referent-honorific	3hon	
referent-humble	3hml	
reflexive	rfl	pronominal or verbal; if the latter, a voice
reinforcement marker	(intns)	recommended term is `intensive'
relative	rel	morpheme marking relative clause, incl. relative pronoun
relative	(rfr)	case; recommended term is `referentive'
remote	(dist)	recommended term is `distal'

remote past	rempst	
repetitive	rep	only if distinct from iterative
reportative	rprt	evidentiality
restricted	sg	in a personal system such as Nandi, Kayapo and other Papuan languages
resultative	res	aspect or aktionsart
reversive	rvrs	
same subject	ss	
secondary object	so	
semelfactive	smlf	
sensory evidential	sens	
separative	(abl)	local case; recommended term is `ablative' until evidence to the contrary
sequential	seq	semantic relation of clause; vs. simultaneous
simultaneous	sim	simultaneity (subordinate clause?)
singular	sg	
singulative	sgt	
sociative	(assoc)	case; recommended term is `associative'
speaker-honorific	1hon	
speaker-humble	1hml	
specific	spec	determination
stative	stat	
subelative	subel	local case
subessive	subess	local case
subject	sbj	verb agreement or cross-reference position
subject verb agreement	idx.sbj	
subjunctive	subj	verbal category typically used in subordinate clauses
sublative	subl	local case
submorphemic unit	0	
subordinator	sr	includes `comp'
superdirective	(supl)	local case; unclear whether needed; recommended term is super-lative

superelative	supel	local case
superessive	supess	local case
superlative	sup	
super-lative	supl	
terminative	term	local case or aktionsart
topic	top	
transformative	trnsf	case: `becoming'; see also `essive'
transitive	tr	
transitive patient	p	only if opposed to both S and A; use OBJ otherwise
transitive subject	a	only if opposed to both S and P; use ERG otherwise
transitivizer	trr	see also `causative' and `extraversive'
translative	trnsl	local case
trial	trl	all trials seen up to now are paucals (Greenberg)
undergoer	ugr	case or verb agreement or cross-reference position for the inactive function in an active system
unrestricted	pl	in a personal system such as Nandi, Kayapo and other Papuan languages
unspecified	unspec	unspecified argument of relational base
validator		recommended terms are `assertive', `declarative', provisionally
venitive	ven	directional
verbalizer	vr	
visible	vs	deixis
visual	vis	evidential
vocative	voc	case
volitional	vol	action carried out volitionally; usually verbal category
volitive	vol	recommended term is `volitional'
1st person	1	
2nd person	2	
3rd person	3	

Back to index

4. Primary linguistic data

4.1. File structure

4.1.1. Database structure

The units of a database are records. A record consists of fields. As to field structure, there are two kinds of databases:

- database with rigid field structure: the field structure is a property of the database and therefore the same for all records;
- database with flexible field structure: the field structure is a property of each record and may therefore differ from record to record.

Database programs usually allow for databases of only one kind. In what follows, a database with flexible field structure will be assumed. However, the flexibility requirements are such that they can be met by a database with rigid field structure simply by providing the maximum number of fields and not filling in some fields in some records.

4.1.2. The text database

The following guidelines apply to a sample text which is used as a self-sufficient set of data. They apply analogously to a set of example sentences that are stored separately and may be retrieved for use in a metalinguistic context.

A database may contain several sample texts (or sets of example sentences). From the title of the text, an abbreviation is derived that is repeated in all of its records (in the record id).

The text is broken down into units whose length does not exceed the size of a print line. Preferably, such a database unit should be a syntactic unit: maximally, a sentence, minimally, a phrase. Each such unit founds a record.

4.1.3. The null record

The first record of a text, called the null record, has the following field structure:

1. **Text id**: abbreviation identifying the text, followed by record number 000.
2. **Title**.
3. **Name of author**.
4. **Date of production**.
5. **Publication**: bibliographical data, if the text has been published.
6. **Type of text**, according to some classification of genres.
7. **Name of analyst**.
8. **Comment**: explanation of any non-standard features, e.g. special symbols.
9. **Date of last modification**.

4.1.4. Body of text database

Each record except the null record has the following maximum field structure:

1. **Record id**: Abbreviation identifying the text, followed by record number (normally three digits, with leading zeros).
2. **Orthographic representation**: Original orthographic form of the text, transliterated if necessary.
3. **Phonetic representation**: Broad phonetic transcription.
4. **Phonemic representation**: Text words represented as sequences of phonemes.
5. **Prosodic representation**: Prosodic structure according to Du Bois et al. 1992.
6. **Morphological representation**: Text broken up into morphemes (each given in morphophonemic representation), with suitable boundary symbols according to section 3.1.
7. **Morphological gloss**: Representation of each item in field 6 by its meaning or grammatical function (grammatical category label), according to section 3.1.
8. **Grammatical tagging**: Grammatical categories of items in field 6, structural information (e.g. bracketing). Details in section 4.2.
9. **Translation**: Idiomatic translation into one of Eurotyp's official languages.
10. **Descriptors**: Keywords identifying interesting linguistic features of the record. They are taken from the set in section 2.2.
11. **Comments**: Free format remarks, including problems.

4.1.5. Subset of fields in a record

The field structure of a given record is an appropriate subset of the above field structure. In the selection of this subset, the following considerations apply:

- #1 is necessary, or else no reference to the record will be possible.
- #2 or #4 or both are necessary.
- The necessity of #3 depends on the kind of text; if the text was never recorded, it will be superfluous.
- The necessity of #4 depends on the language. It will be necessary if it differs substantially from #6 minus morphological boundary symbols, as is the case in languages with a heavily obliterative phonology.
- #6 is necessary for a language with some morphology and if the data are to be used for grammatical analysis.
- #7 is necessary if the data are to be used for grammatical analysis.
- #9 is necessary if the data are to be used in arbitrary scientific contexts. It will generally be in English.
- The records of a text should be uniform in terms of fields #1 - 9. #10 and 11 are optional from record to record.

Du Bois, John W. & Schuetze-Coburn, Stephan & Paolino, Danae & Cumming, Susanna 1992, *Discourse transcription.* Santa Barbara: UCSB, Department of Linguistics (Santa Barbara Papers in Linguistics, 4).

4.2. Tagging: Coding the linguistic structure of a text

[to be filled in]

Back to index

5. Typological Questionnaires

5.1. General observations and guidelines

A typological questionnaire will in the following be understood to be any questionnaire which is used with the aim of collecting parallel information about the members of a set of languages. The information one is trying to collect may be of two basic kinds corresponding to the two following types of question:

1. How do you say 'X' in L?

2. How does the phenomenon P function in L?

A questionnaire may aim at either of these, or both. Various terms have been used for the 'pure' types of questionnaires, such as:

a. translation questionnaires
   elicitation questionnaires
   primary data questionnaires

b. question questionnaires
   analytical questionnaires

Some of the terms here are not optimal: 'translation questionnaires' because it excludes other methods of eliciting primary data, 'question questionnaires' mainly because it sounds tautological. We recommend the terms elicitation questionnaires and analytical questionnaires for the pure types and mixed questionnaires for the rest. In constructing a typological questionnaire of whatever type several general points should be considered.

Data pertaining to the informant. The following information about the informant should be elicited: name, contact address, sex, profession, degree of competence in the language under investigation, languages spoken by the informant other than the one under investigation.

Contents of the questionnaire. A table of the contents of the questionnaire should be provided and the type of questions featuring in the questionnaire (yes/no, if-, multiple choice, open, example sentences) should be made explicit. Unambiguous instructions should be given on how to respond to each type of question, including how to indicate lack of information.

Structure of the questionnaire. The questions in the questionnaire should be divided into numbered sections. Each question should have a unique number specifying both the relevant section and the consecutive number of the question. In structuring the questionnaire and formulating the questions attention should be given to reducing the burden of the task of the informant by:

> a. clearly stating the point of the question and the level of specificity of the answer required;

b. exemplifying controversial or ambiguous terms;

c. informing about the existence of further more detailed or related questions;

d. providing potential answers to the questions;

e. making it clear that the question is applicable only to a particular subset of languages, if this is the case.

To reduce error, check-questions should be included. The informant should be provided with the opportunity of making additional comments.

5.2. Elicitation questionnaires

Although elicitation questionnaires have been used by linguists for a long time, it is only recently that attention has been paid to the methodological problems connected with their use in typology. We give here a number of recommendations which can guide the construction and use of elicitation questionnaires within Eurotyp.

Choice of method. Generally, the fastest and most convenient way of collecting large sets of data from a number of different languages is the translation method – where a set of sentences or texts are translated from a source language into the languages under investigation. We list below some advantages and drawbacks of this method:

Advantages of the translation method:

- It is probably the safest way of getting a speaker to produce utterances which correspond closely in content and structure to the intention.
- It is fast and simple.
- It results in written output which is easy to analyze.
- It does not, in general, demand any knowledge of grammar or linguistics on the part of the informant.

Drawbacks of the translation method:

- It presupposes good knowledge of the source language, or else, the help of intermediary persons whose influence on the final product is unpredictable.
- Certain types of phenomena are hard to elicit: a) marked syntactic constructions (like passives), b) secondary readings of lexical and grammatical items, c) informal and substandard forms.
- The form of the source sentences may influence the output in unpredictable ways.

To avoid some of the drawbacks, it is often recommendable to use additional methods. These will often be of the kind 'Is such-and-such a sentence acceptable in such-and-such a situation?' or 'Can such-and-such an utterance have this additional reading?'. It is crucial when constructing such questions to consider to what extent they will presuppose theoretical knowledge.

Choice of informants. In general, it is wise to enforce a 'native speaker requirement' on informants. Experience with informants with second-language competence has not been encouraging, at least not in areas such as tense and aspect. What is a native

speaker? This is not an easy question to answer. An operational definition might be 'a person who has used the language actively and daily when growing up'. The native speaker requirement is a necessary but not sufficient condition: experience shows that staying away from the area where the language is spoken for prolonged periods may have rather disastrous consequences for the knowledge of a language, even if it was once 'native'. Regrettably, for many languages, the majority of the easily obtainable informants may be persons who have spent a long time in foreign environments.

Experts vs. informants. Even if translation does not in principle require knowledge of grammar or linguistic theory, there are many situations where an 'expert' is needed in addition to the informant. This is of course in particular the case if the questionnaire is of a mixed kind. An expert is also needed if the informant is illiterate or does not have sufficient knowledge of the source language. In these cases, it is sometimes necessary to make intermediary translations of the questionnaire into some language known to the informant. This, as was noted above, may introduce unpredictable noise into the translation process.

How many informants? It is of course an advantage if one can get data from more than one informant for each language, although in many cases, this may not be feasible. A larger number of informants decreases the risk for errors and makes it possible to get an idea about variations in usage. When administering a questionnaire to a large set of speakers of one language, it may however be more practical to do in the form of a completion rather than a translation task, i.e. to give them sentences in their own language with left-out words to fill in or alternatives to choose between. As in the case with intermediary translations, this may introduce noise into the system, since the initial translation that is necessary to transform the original questionnaire may distort it in ways not intended by its constructor. This, then, is a step that demands special caution.

Keep informants separate in the data-base. When there are several informants from one language, the person who handles the data may feel tempted to integrate the responses from the different informants into one data-set. This should be avoided,. since it makes it more difficult to study variation and interdependencies between individual responses to different questions. The responses of individual informants should therefore be kept separate in the data-base.

The importance of the context. It is often the case that isolated sentences allow of multiple translations. It is therefore essential to indicate the (extralinguistic and linguistic) context in which the utterance is supposed to be made. In particular, care should be taken to see to it that there is enough information for the correct choice of grammatical categories such as tense, mood, and aspect. (This goes also for investigations that in themselves have nothing directly to do with those categories!)

Avoid 'grammar book' examples. Try to construct sentences that could really be used in real life and avoid examples that too directly reveal their grammar book origin.

The cultural bias problem. A novice constructor of typological questionnaires is usually not prepared for the difficulties connected with making up examples that are not culturally or geographically biased. Some examples.

> It seems to be considerably easier to find linguistic universals than to find universal food and drink items. Thus, neither beer, wine, water or milk is a natural

> direct object for the verb 'drink' in all languages - alcohol is taboo in many places, water may not be considered a worthy substance to drink at all, and a large part of the earth's population just cannot digest milk.
>
> Linguistic examples are often criticized for being sexist. However, the typologist who tries to construct politically correct examples will soon find that informants refuse to translate them 'because men/women don't do that sort of thing'.
>
> Certain things may be entirely taboo, for various kinds of reasons. A verb like 'die', for instance, is problematic: it may well turn out that one has to use a paraphrase, either because it is unrespectful to say of someone that he died, or because it is considered dangerous to mention death.

The real conflict is between striving to avoid cultural bias and trying to construct natural, everyday utterances. The trouble is that everyday life is heavily culture-bound. The only way of cutting the Gordian knot is to allow for culturally bound items to be replaced by others in the translations. The following instructions are taken from the Future Time Reference Questionnaire used in Theme Group 6:

> A word or phrase in English may have no natural equivalent in L, or it may be felt that the sentence describes a situation which is foreign to the culture in which L is spoken. In such cases, try to find an analogous word, phrase, or sentence. In doing so, try to choose concepts from the same general area and keep as close to the grammatical structure of the original as possible. For instance, for 'in the forest', 'in the desert' is a possible substitute. If the original says 'write a letter', choose e.g. 'bake a cake', 'build a hut', 'make a net'.

This is a point where an expert may have to intervene, which makes the strategy harder to apply.

Requirements on data. The data should be of such a quality that persons who are not familiar with the language in question are able to make use of them and that example sentences can be used in linguistic reports. For this reason, it is highly recommended that all primary data be equipped with inter-linear glossing when entered into the data-base. (Needless to say, this is not something you could demand from a layman informant.) Also, avoid hand-written data: such texts in languages that you do not know are notoriously difficult to decipher. Ideally, of course, data should be converted into computer files as early as possible in the process.

Format of data. If data are to be entered into a data-base, care should be taken that submitted files are in a format suitable for being imported into the data-base in question. Since this depends on the software we refrain from giving proposal for a standard here.

5.3. Analytical questionnaires

One of the ways of cross-checking the correctness of the responses to analytical questions is by requiring the informant to provide relevant language examples. This being the case, most of the issues pertaining to the elicitation of appropriate language examples discussed in connection with elicitation questionnaires pertain also to analytical questionnaires.

Analytical questionnaires are more theory-bound than elicitation questionnaires. Needless to say, the more theory-specific the questions are the more restricted the class of the potential informants, and the smaller the chance of ensuring a good coverage of languages. The general recommendation is, therefore, to try and formulate questions in as theory-neutral terms as possible.

Type of questions. Of the various type of questions that analytical questionnaires may contain, the multiple choice question (with an option 'other') places the smallest burden on the informant and the largest on the author of the questionnaire. Such questions are the most likely to elicit answers and are also the easiest to process. Multiple choice questions are often combined with an *if*-question. This, on the whole, should be avoided and rather substituted by providing a preceding *Yes/No* question. Otherwise, if the informant is not in a position to answer the multiple choice question, which may well be the case, the analyst does not know whether the phenomenon occurs in the language or not.

5.4. Survey of questionnaires used in the Eurotyp project

So far twenty seven questionnaires have been used in the Eurotyp project. These are listed below according to which of the nine thematic groups they have been constructed by.

Group 2: Constituent Order

Questionnaire 1: Word order
Type: Analytical
Author: Anna Siewierska

Questionnaire 2: SOV-order in SVO-languages, complementizers and word order
Type: Elicitation
Author: Anders Holmberg

Questionnaire 3: Discourse configurationality
Type: Analytical
Author: Katalin Kiss

Group 3: Subordination and Complementation

Questionnaire 1: C and I systems
Type: Analytical
Author: Ian Roberts

Questionnaire 2: Complement types
Type: Elicitation
Author: Karina Vlaming

Group 4: Actancy and Valency

Questionnaire 1: Actancy
Type: Analytical
Author: Gilbert Lazard/Jack Feuillet

Group 5: Adverbial relations, operators and connectives

Questionnaire 1: Adverbs and particles of change and continuation
Type: Mixed
Author: Johan van der Auwera

Questionnaire 2: Adverbial conjunctions
Type: Mixed
Author: Bernd Kortmann

Questionnaire 3: Adverbial quantification
Type: Mixed
Author: Juan Carlos Moreno

Questionnaire 4: The internal structure of adverbial clause
Type: Mixed
Author: Kees Hengeveld

Questionnaire 5: Equality and similarity
Type: Mixed
Author: Oda Buchholz and Martin Haspelmath

Questionnaire 6: Concessive conditionals
Type: Mixed
Author: Martin Haspelmath and Ekkehard König

Questionnaire 7: Sentence adverbs
Type: Elicitation
Author: Paolo Ramat and Davide Ricca

Questionnaire 8: Converbs
Type: Analytical
Author: Igor' Nedjalkov

Group 6: Tense and aspect

Questionnaire 1: Future time reference
Type: Elicitation
Author: Östen Dahl

Questionnaire 2: Perfect
Type: Mixed
Responsible: Jouko Lindstedt

Questionnaire 3: Progressive
Type: Mixed
Responsible: Pier Marco Bertinetto

Questionnaire 4: Absentive
Type: Mixed
Responsible: Casper de Groot

Group 7: Noun Phrase Structure

Questionnaire 1: Number
Type: Analytical
Author: Greville Corbertt

Questionnaire 2: Gender
Type: Analytical
Author: Greville Corbertt

Questionnaire 3: Universal quantification
Type: Mixed
Author: David Gil

Questionnaire 4: The configurational count/mass typology
Type: Analytical
Author: David Gil

Questionnaire 5: Numerals
Type: Analytical
Author: James Hurford

Questionnaire 6: Nominalizations
Type: Analytical
Author: Maria Koptjevskaja-Tamm

Questionnaire 7: Genitives
Type: Mixed
Author: Maria Koptjevskaja-Tamm

Questionnaire 8: Descriptors of NP internal structure (Working Paper 4)
Type: Analytical
Author: Edith Moravcsik

Questionnaire 9: The dual
Type: Analytical
Author: Frans Plank

Questionnaire 10: Derivation and Inflection (Working Paper 10)
Type: Analytical
Author: Frans Plank

Questionnaire 11: Ellipse and inflection of determiners and modifers in coordinate NPs (Working Paper 11)
Type: Analytical
Author: Frans Plank

Questionnaire 12: Co-occurrence of possessives with articles and demonstratives (Working Paper 11)
Type: Analytical
Author: Frans Plank

Questionnaire 13: Cases and appositions (Working Paper 13)
Type: Analytical
Author: Frans Plank

Group 8: Clitics and non-lexical categories

Questionnaire 1: Clitics
Type: Analytical
Author: Riet Vos

Back to index

6. Coding, storage, retrieval and analysis of analytical linguistic data

In what follows the term analytical data will be used to denote all systematic and linguistically relevant observations on linguistic entities, such as words, clauses or stretches of discourse, for a specific language, on any level of description, either in terms of a specific theory or in terms of a more general descriptive framework. Some examples of such data are:

a. the basic order in main and subordinate clauses
b. the way different tenses are expressed
c. the position of the complimentizer in relation to the clausal category it selects
d. the occurrence and the type of vowel harmony
e. the occurrence of pronominal politeness forms and whether it is marked or unmarked in spoken language.

Much work in language typology is based on collections of this type of data. Often they are gathered by means of a questionnaire such as the ones described in section 5 of the guidelines. Even before a questionnaire is constructed, the researcher should establish a (tentative) model of the linguistic domain concerned. From such a model, the actual questions should be derived, and structured. The questionnaire results may then be interpreted in terms of the initial model. This may eventually lead to modifications and refinements in the model. If no such modelling had been done before the questionnaire was constructed, possibly since the knowledge of the domain was still too sketchy, the modelling may be done precisely on the basis of the incoming questionnaire results.

If the properties of the model, composed prior to the questionnaire or based on its results, comply to certain minimal formal standards, then the incoming data may be formalized as well, and processed by a computer. For this, standard data base and data analysis programs may be used. Well-known examples of the former are D-Base, Oracle and Shoebox. Section 10 of the guidelines gives a standard format. Data analysis may be performed with statistical packages such as SPSS. The drawback of these standard programs is that they provide only very general facilities, and are not tuned to the more specific needs of the linguist. The data formats that are required are often quite 'unnatural', i.e. not of the form in which observations often are formulated by linguists. A convenient means for doing so is presented below.

First, section 6.1 discusses the features of a theory-independent system for representing analytical data. And in section 6.2 these features are implemented in a formal descriptive language called Domain Structuring Language (DSL), that we think provides a more natural way for linguistic data representation than the formalisms generally used in data processing. A set of computer programs, based on DSL, has been devised that may be used for coding, storing and analyzing such data, taking over part of the work that has to be done by the linguist. They implement a number of data analysis techniques that may not be found among the standard software packages. For standard tasks, interfaces have been built in that give access to some of the generally available packages. The programs are briefly described in section 10.1. A full

description of the possibilities is given in working papers 2-3 and 2-5 of the Theme Group on Constituent Order.

6.1. Coding a linguistic domain

In empirical sciences, a descriptive domain is often structured by way of variables and values. Variables provide a syntagmatic structuring to the domain, i.e. they give the dimensions that exist in parallel. Any variable has a set of values assigned to it, that give a paradigmatic structuring to the domain. In the simplest form, all variables are thought to be relevant for all cases, i.e. all entities that are the object of analysis, such as languages for linguistics. Variable values are mutually exclusive, i.e. any case selects precisely one value for each variable. A schematic version of this representation scheme is the data matrix of figure 1, with the cases on the vertical and the variables on the horizontal dimension. Every cell contains precisely one value, i.e. the score for the given case in the row in regard to the variable in the column.

	variable 1	variable 2	variable 3
case 1	value 1_1	value 1_2	value 1_3
case 2	value 2_1	value 2_2	value 2_3
case 3	value 3_1	value 3_2	value 3_3

Figure 1. Data matrix

In order to ensure that such a matrix will be complete, special values may be employed to code the fact that, for a specific case, no value is known for some variable, or that some variable is not relevant altogether. These are commonly called 'missing values', and may receive special treatment under analysis. Although, formally, this data matrix representation method is rich enough to code any type of analytical data, it does so in a manner that diverges from the way linguistic observations are normally formulated.

To enable the representation of language facts in a way closely corresponding to the actual practice of linguists, the above simplified variable - value format has been enriched by several conventions. Each of these will be discussed briefly in turn.

a. Multiple values

The canonical way of coding language phenomena the realization of which either typically does or may involve a set of options is to set up a separate variable for each option and supply it with a yes/no value.

Needless to say, this can be very cumbersome. When dealing with a phenomenon with numerous options, for example, the type of case distinctions made, such a solution would require over 40 variables.

What is therefore necessary is to allow variables to be associated with multiple values

ranging from two to whatever number is required.

b. Structured values

It is generally impossible to predict all the values for all the variables in terms of which one has structured a given domain. It is therefore advantageous to use values which can be potentially concatenated to form new values. Moreover, if some of the values used have internal structure, i.e. they are decomposable, it may be possible to form new values from the composite parts of the old. This may considerably facilitate the drawing of generalizations over values. For example, given the three values SOV, OSV and OVS each of which can be decomposed into S, O and V, we can generalize over OV (all 3 of the values), SV (2) or OS order (2).

c. Scales

Often, linguistic data, and more precisely values for variables, may be subject to scaling along some dimmension. For example, in a number of languages the location of adverbials of setting is typically sentence initial or sentence final. Clause internal placement would be considered as marked. This phenomenon could be coded in several ways. We could have separate variables for the unmarked and the marked location. We could have one variable and then concatenate 'marked' and 'unmarked' to the location values. We extended the variable-value scheme with the possibility of associating the values of particular variables with additional scalar values. The above example on adverbials may then be coded by providing a markedness scale for the variable 'location of adverbial of setting' and allowing one to specify which of the values chosen for this variable such as 'initial', 'internal' and 'final' are marked or unmarked.

d. Conditions

In a variable-value scheme, we may have to deal with two types of conditions: syntagmatic and paradigmatic ones. Syntagmatic conditions occur when a language has several values for one variable, each occurring under different linguistic circumstances. For example, the relationship between the possessor and possessed in English may be coded either by a preposition between the two or by a genitive enclitic on the possessor. There are no specific (categorical) conditions on the use of the first construction. The use of the latter, however, is restricted to animate nouns that have to be relatively short. Instead of coding such conditions in the values, they can be specified as a condition on a value.

Paradigmatic conditions involve dependencies between values. It is sometimes the case that the relevance of some variable for a language is dependent on the value for some other variable. Cf. when a language has no articles, variables coding their precise aspects, such as their formand location, are irrelevant for that language. To capture this fact it is desirable to have a variable referring to the existence of articles and to place this variable before the ones pertaining to specific features of articles or their use. Thus if the first variable receives the value "no" all the dependent variables will be defined as irrelevant.

These extensions lead to a four-dimensional way of structuring a linguistic domain, in terms of variables, values, scales and conditions.A schematic representation may look as follows:

Figure 2. Graphic representation of a linguistic domain (to be filled in)

6.2. The representational language

The DSL language provides the means of structuring and coding a linguistic domain precisely in terms of the four dimensions mentioned above. A complete formal definition is found in appendix A. Here, only some representative examples of its use for the coding of information will be given. The information stems from the domain of word order, case systems and adverbials. The overall structure of a domain definition is as follows:

(1) a. Domain description
b. Scales
c. Missing values
d. Condition variables
e. Domain variable descriptors

(1a) is a short text describing the domain, meant for identification purposes only. (1b) are the definitions of the relevant scales. (1c) provides a list of those values that should be considered as 'missing' for any of the variables in the domain. The condition variables (1d) do not necessarily belong to the domain proper, and may be used in syntagmatic conditions. For the word order data questionnaire, (1a) through (1d) look as follows (abbreviated):

(2) a. Word Order Data, Version 1.1 Nov92
b. Scales: Preference=(oblig, pref, non_pref, imposs)
Restrictive=(fav, restr)
c. Missing: (not_known, not_clear, not_present)
d. External: Position1=(expletive, adverb)
DefSubj=(def, indef)

Variable Position1 in the list of externals may be used in syntagmatic conditions to represent what is in the first position; variable DefSubj will represent the definiteness status of the subject.

The basic entities of DSL are the domain variable descriptors (1e), that give a formal description of the respective variables that paradigmatically code the domain concerned. They are provided as an ordered list of individual descriptors. A variable descriptor is built up in the following way:

(3) a. Variable name
b. Reference
c. Variable label
d. Parameters
e. Value set

The variable name is a short, unique indication for the variable for quick reference, not necessarily mnemonic (cf. Var207). The reference is an (optional) indication of the source of the value of the variable, such as a particular set of questions in a questionnaire. It is basically a convenience for checking the original source of the data

in a questionnaire. The variable label is a description of the meaning of the variable, for instance, 'order of the possessor and possessed', 'existence of impersonal passive', etc. Parameters characterize the variable in terms of its type (e.g. multi-valued; complex values), associated scales, and conditions. The value set provides an exhaustive enumeration of the values for the variable. Some of the above features are illustrated in (4):

(4) V2.1 (Q2.1 - Q2.4)
Main bare intransitive orders
parameters= (multi=2, scale=Preference, syntagmatic_condition)

a. SiV
b. VSi

The parameters given here should be interpreted as follows. 'multi=2' means that to a maximum of two values (in this case in fact: all) may be chosen from the list of predefined values. Any one of them may be associated with a value from the scale 'preference'(see (2b)). The presence of the 'syntagmatic_condition' parameter allows a syntagmatic condition to be stated for any value chosen for the variable for some language. Definition (4) gives the possible values of domain variable V2.1 along all relevant dimensions. For a particular language that is in the sample, the values for a variable are coded according to the following syntax:

(5) var = value_1 - [(scale=value) , (syntagm_condition)] /
value_2 - [(scale=value) , (syntagm_condition)] /
...

When scale and condition are irrelevant the part after the hyphen may be left out. Syntagmatic conditions are boolean expressions over variables and values from the set of condition variables. In conditions, the operators AND and OR may be used, as well as unlimited embedding by means of brackets. In the simple expressions, both the equality operator = and the unequality operator # may be used. We may end up with the following value description for variable V2.1 for some specific language L. In other words: these are the contents of the cell on row L and column V2.1 of the data matrix:

(6) V2.1= VSi -[(preference=obligatory) ,
((Pos1=expletive) OR (Pos1=adverb) OR (DefS=indef))] /
SiV

This should be interpreted as: in intransitives of L the subject follows the main verb as the obligatory order in case there is an expletive or adverb in the first position and also in case the subject is indefinite. In all other cases the order is subject-verb. An example of a paradigmatic condition is given in (7). Note that, while syntagmatic conditions are part of the data for a specific language, paradigmatic conditions are on variable descriptors:

(7) V29.8a (Q29.3 - Q29.4)
Nominal cases
parameters= (multi=10, paradigm_condition=[V29.8a=none > V29.9])

Thus, if V29.8a has value 'none' for some language, the variable immediately dependent on it is V29.9, not the default one, i.e. the variable of which the descriptor immediately follows that of V29.8a, say V29.8b. This compares to the relationship between variables V1, V2 and V3 in figure 2 above. In this type of paradigmatic condition, before the 'greater than' sign we have a simple equality or unequality expression for the current variable. After the sign we find a variable label. There may be more than one such paradigmatic condition in a variable descriptor.

A second type of paradigmatic condition on variable V consists of an implication with a boolean expression over variables higher than V for its premise and a subset of the values of V for its conclusion. Assuming that variables V7 and V8 are higher than V9, i.e. their descriptors come before that of V9, we could have the following for paradigmatic condition V9:

(8) paradigm_condition=[(V7 = val7a AND V8 # val8c) -> (val9b, val9d)]

meaning that if V7 has value val7a and V8 does not have value val8c then the selectable values for V9 should be restricted to the subset val9b and val9d. The latter version of the paradigmatic condition could be called an input condition, the former one an output condition: For any language, the former determines the options for a variable before it is assessed, in the context of the values determined for higher variables. The latter determines the configuration of variables lower than some variable after its value has been determined. A last option in the parameter field of a variable descriptor is the 'open' option. This signals that, apart from the predetermined set of values, the variable may be assigned any other value that has not been specified in advance.

Back to index

7. Alphabets and character sets

7.1. Transliteration/transcription of languages with Non-Latin scripts

7.1.1. Transliteration/transcription

All material from languages that are written in a writing system other than the Latin alphabet must be transliterated/transcribed.

As a rule, one transliterates when the original spelling reflects the pronunciation fairly closely, and one transcribes when the orignal spelling does not represent the sounds well. In Europe, only Modern Greek is generally transcribed. All other languages that use non-Latin writing systems are transliterated.

7.1.2. The Cyrillic alphabet

7.1.2.1. The Slavic languages

Six modern Slavic languages are written in a version of the Cyrillic alphabet: Russian, Ukrainian, Belorussian, Bulgarian, Macedonian, and eastern Serbo-Croatian (as well as Old Church Slavonic, which is written in Cyrillic in most textbooks and editions).

The following letters are common to all Cyrillic alphabets and are uniformly transcribed: note

а	a	о	o
б	b	п	p
в	v	р	r
д	d	с	s
е	e	т	t
ж	z	у	u
з	z	ф	f
к	k	ц	c
л	l	ч	č
м	m	ш	š
н	n		

Each language has some special letters that do not exist in all languages or are transliterated differently in different languages.

Russian

г	g	ъ	"
ё	ë	ы	y
и	i	ь	′
й	j	э	è

х	x	ю	ju
щ	šč	я	ja

Ukrainian

		є	ji
г	h	й	j
ґ	g	х	x
ѓ	je	щ	šč
и	y	ю	ju
і	i	я	ja

Belorussian

		ы	y
г	h	ь	´
і	i	э	è
й	i	ю	ju
ў	w	я	ja
х	x		

Bulgarian

		щ	št
г	g	ъ	â, ă
и	i	ю	ju
й	j	я	ja
х	x		

Serbo-Croatian (eastern)

		љ	lj
г	g	њ	nj
ћ	ć	х	h
и	i	џ	dž
ј	j	ђ	đ

Macedonian

		ѕ	dz
г	g	и	i
ѓ	ǵ	ј	j
ќ	ć	х	h
љ	lj	џ	dž
њ	nj		

Old Church Slavonic

		лъ	ĭ
г	g	ь	ĕ
s	dz	ю	ju
и	i	ѣ	ja
і	i	є	je
ђ	ǵ	ѧ	ę
оу	u	ѫ	ǫ
х	x	ѩ	ję
щ	št	ѭ	jǫ
ъ	ŭ		
ы	y		

7.1.2.2. Non-Slavic languages

Many non-Slavic languages spoken in Russia and other parts of the former Soviet Union use the Cyrillic alphabet. No transliteration conventions for these languages are given here because

(a) most of these languages are not widely studied outside of Russia and the former Soviet Union

(b) some of them (especially the languages of Central Asia) are in the process of shifting to other alphabets, especially the Latin and the Arabic alphabets

7.1.3. The Greek alphabet

The Greek alphabet is used for Ancient Greek and for Modern Greek. For Ancient Greek, a transliteration is used because the spelling reflected the pronunciation fairly closely. For Modern Greek, linguists usually use a transcription because many spelling conventions from Ancient Greek are still used in Modern Greek spelling.

7.1.3.1. Ancient Greek

Cf. Martinet, André. 1953. "A project of transliteration of Classical Greek." *Word* 9.2: 152-161.

α	a	ν	n
β	b	ξ	x
γ	g	ο	o
δ	d	π	p
ε	e	ρ	r
ζ	z	σ	s
η	ē	τ	t

θ	th	υ	u
ι	i	φ	ph
κ	k	χ	kh
λ	l	ψ	ps
μ	m	ω	ō

Special conventions:

(a) accent marks and trema are written as in Greek,

e.g. ά = á, έ = è, ΐ = ĩ, αϋ = aü
αί = aí, εῦ = eũ

(b) spiritus asper is transliterated as *h*, e.g. ἁ = ha

(c) iota subscriptum is not subscript, e.g. ῃ = ēi, ῳ = ōi; ᾳ is āi

(d) γ before nasal stop can be transliterated as *n*, e.g. ἄγγελος= ángelos

7.1.3.2. Modern Greek

Cf. Joseph, Brian & Philippaki-Warburton, Irene. 1987. *Modern Greek*. London: Routledge.

(a) Vowel inventory: /i e a o u/, transcribed as in IPA.

(b) Consonant inventory:

/p	t		c	k
b	d			g
f	θ	s	ç	x
v	ð	z	j	γ
m		n		
		l	ʎ	
		r/		

(c) The sixteen consonants /p, t, k, b, d, g, f, s, x, v, z, j, m, n, l, r/ are transcribed as in IPA.

(d) The fricatives /θ, ð, γ/ may be transcribed as in IPA; or alternatively they may be transcribed by the digraphs <th>, <dh>, <gh> for typographic convenience.

(e) The palatal sonorants /ɲ, ʎ/ are always transcribed by the digraphs <nj>, <lj>.

(f) The palatal obstruents (c, ɟ, ç/ contrast with /k, g, x/ only before back vowels, so they need to be distinguished from these only in this environment; before front vowels, dorsal obstruents are always palatal. Thus,

και	/ce/	<ke>

κιόλας /colas/ <kjólas>

before back vowels:

/c/ = <kj>

/ɟ/ = <gj>

/ç/ = <xj>

before front vowels:

/c/ = <k>

/ɟ/ = <g>

/ç/ = <x>

Stress should also be indicated in Modern Greek transcriptions.

7.1.4. The Hebrew alphabet

Among European languages that are widely studied, only Yiddish is written in the Hebrew alphabet. The transliteration of YIVO (New York center for Yiddish studies) should be used.

א	a	ײ	ey
ב	b	ײ	ay
ג	g	כ,ך	kh
ד	d	ל	l
ה	h	מ,ם	m
ו	u	נ,ן	n
וי	oy	ס	s
וו	v	ע	e
ז	z	פ	p
זש	zh	פ,ף	f
ח	(Hebrew)	צ,ץ	ts
ט	t	ק	k
טש	tsh	ר	r
י	i	ש	sh
יי	y	ת	(Hebrew)

Two letters are only used in Hebrew loanwords. Note that Hebrew loanwords are not transliterated, but transcribed.

7.1.5. Others

7.1.5.1.

For the Armenian alphabet, see

> Minassian, Martiros. 1980. *Grammaire d'Arménien oriental*. Delmar, N.Y.: Caravan Books.

7.1.5.2.

For the Georgian alphabet, see

Vogt, Hans. 1971. *Grammaire de la language géorgienne*. Oslo.

For both Armenian and Georgian, see also

> Comrie, Bernard. 1981. *The languages of the Soviet Union*. Cambridge: Cambridge University Press, p. 288-289.

7.2. Survey of systems in use

Rendering other languages than English (and possibly Latin and Dutch) has been a problem as long as computers have existed, and it has been particularly acute for linguists, who more often than not need to give examples from several languages within one text. Gradually, computer hardware and software has become more suited to fulfil these needs, but at present, we are still far from a general and well-working solution. At least two proposals exist for a 'universal character set' which would comprise virtually all characters needed to render the languages of the world and in addition, the special characters used e.g. in phonetic transcriptions and mathematical and logical formulae. The most promising one at present seems to be Unicode, a two-byte encoding system for characters developed by an international consortium (Unicode Inc., 1965 Charleston Rd, Mountain View, CA 94043, USA.) Representing each character as two bytes means theoretically that there is room for 65,536 characters, which makes it possible to include not only the Roman alphabet and its extensions but also e.g. the essential parts of the Han characters that form the basis for the Chinese, Japanese and Korean writing systems. The fact that major software companies are members of the Unicode consortium gives good hope that it will be adopted in the future. However, this is of little help in the present situation: virtually all existing software packages build on one-byte representations of characters, which makes it impossible to exploit Unicode's principles. (For details on Unicode, see Sheldon 1991).

The development of word-processing systems has now got so far that representing West European languages is usually no problem, and files can also relatively easily be transferred between the major word-processors and between the IBM and Macintosh worlds, provided that you follow the instructions in the manuals. Also, it is usually possible, with varying degrees of difficulty, to render at least the more common diacritics and special characters in the major word-processors. Within Eurotyp, problems arise above all when files containing data in 'non-EC languages' are exchanged between different participants, or when other software (such as data-base programs) are used which does not provide for special characters. The following solutions are recommended:

1. If possible, the persons involved should agree on one word-processing system.

2. As a second alternative, there should be an agreement on a simple way of representing characters that are not in the 7-bit Ascii character set. This ensures that data can be transferred without loss of information between virtually any two computers and by virtually any channel of transmission, and that it can be entered using any existing text editor. Below, an example of such a standard is given; it has been used in Eurotyp Theme Group 6 and seems to work fairly well for most purposes.

A general principle is also to choose representations which do not contain more 'fancy' features than necessary. Standard orthography should be used when possible; languages which use non-Roman writing systems should be transliterated rather than transcribed phonetically.

7.3. Coding diacritics and special characters

Diacritics and special characters not found in the standard or extended ascii character set should be rendered as in Table 1. *x* stands for any character. Example: *šāh* should be entered as s\a5h. Other special characters are given in Table 2.

Notice the following: The Ascii code of # is 35, that of $ is 36. If you are using a 7-byte system, please check that you are using the correct characters, since some national standards have other characters in those places. Avoid the national accented characters if you are not certain that the end-user can identify the standard you have been using.

Table 1.Codes for diacritics

ẋ	x1
x́	x2
x̂	x3
x̃	x4
x̄	x5
x̌	x6
ẋ	x7
ẍ	x8
x̲	x9
x̊	x0

Table 2.Codes for other special characters

ð	d#
þ	t#
ʔ	?#
ɔ	c#
ə	e#
ɬ	l#
ŋ	n#

x̰	x$
x̥	x%
x̋	x&
x̯	x*
x̆	x\

Reference

Sheldon, Kenneth M. (1991). "Ascii goes global." *Byte*, July 1991, 108-116.

Back to index

Note: if you cannot see the symbols download Arial Unicode MS place it in your font folder (please note that the file is BIG > 15 MB)

8. Style Sheet for Authors and Editors Preparing Conversion Copy

8.1. The manuscript preparation on disk

The volume editor informs the authors of the programs to be used in preparing the disks, ensures that they are in possession of this style sheet, and distributes to them the publisher's most recent technical instructions on disk preparation. The author sends to the volume editor a copy of his or her text on disk, together with the corresponding printout. The volume editor submits the material for the volume to Mouton de Gruyter, preferably all on one disk. The publisher has the disk or disks checked for their convertibility and also checks the texts to make sure that there are no deviations from the style sheet. The publisher will inform the volume editor of any departures from the style sheet that need to be rectified and of any further requirements for the formatting of the disks.

The editor sends the final edited version on disk, together with the corresponding printout, to the publisher. This version must be complete and correct in every way, since corrections made later on are extremely costly and must be kept to an absolute minimum. The publisher now prepares the publication schedule.

The authors receive the galley proofs of their own articles so that they can check that no parts of the text have become lost during the conversion process, that the figures have been correctly mounted, and that everything is in order. The authors select the keywords to appear in the index, compile a list of these words, and mark in the proofs (for example, with a coloured highlighter pen) each occurrence to be indexed. The author then returns the corrected galley proofs to the volume editor, who collates the corrections with his or her own and returns them to the publishers.

The page proofs are sent to the volume editor only, for final approval. He or she prepares the manuscript for the index, using the lists of keywords provided by the authors and entering the corresponding page numbers. It will save time at the page proof stage if the "skeleton" of the index has been prepared beforehand and only the page numbers remain to be added.

8.2. Ensuring consistency

Editors must make sure that every disk is prepared identically as to format. If different people are handling the input, consistency must be guaranteed.

Items to be checked for consistency throughout the printout include the use of italic typeface for words to appear in italics, superscript figures for note numbers in the text, the correct representation of hyphens and dashes (see 8.7f), one line-space before and after block quotations, etc.

8.3. The manuscripts

a) In general, the text begins flush left throughout. Exceptions are the beginning of each paragraph except those immediately after headings, and block quotations.

b) Do not use right-margin justification.

c) Do not break words at the ends of lines.

d) The text must be run on. This means that there are no "hard" carriage returns except at the ends of paragraphs.

e) Lists should not be indented.

f) Never use the space bar to indicate indentation for extracts, examples, etc. Always use tabs or a paragraph code. (While the numbers of the examples should always appear flush left, you will need to use the tab to line up the texts of the examples themselves.)

8.4. Titles and headings

a) The text should be divided into numbered sections and, if necessary, subsections, with appropriate headings. For all headings please use the following numbering system:

1. Main heading
1.1. Section heading
1.1.1. Subsection heading

b) All headings, including chapter titles in the text and in the table of contents, begin flush left.

c) Do not italicize titles or headings with the exception of words or phrases in them which are to be printed in italics (see section 8.6).

d) Do not end a title or heading with a period when it is to be set on a line separate from the text.

e) The first line of text following a heading or subheading should start flush left (not indented); all subsequent new paragraphs should be indented with the tab. Do not put a blank line between paragraphs in the same section.

f) Capitalize only the first letter of the first word and of other words which the orthography of the languages requires to begin with a capital letter (e.g., proper nouns). This also applies to the table of contents, and to titles of books or articles cited in the text.

8.5. Quotations

a) Short quotations (no more than sixty words) should be run on (i.e., included within the text) and enclosed within double quotation marks.

b) Longer quotations (more than sixty words) should appear as a separate block, indented left and right, and separated from the text by an extra line space above and below. They are not to be enclosed within quotation marks. Use generic codes to mark the beginning and end of indentations for block quotes.

c) All quotations should follow the original text exactly in wording, spelling, and punctuation. Any additions by the author should be indicated by square brackets. Indicate omissions by ellipsis points without brackets.

d) Material quoted from works in languages other than English should be given first in the original language, followed by the translation enclosed in square brackets. (Please see section 8.9 for treatment of examples in languages other than English.)

8.6. Italics and emphasis

a) Foreign-language expressions that have not become standard in English should be italicized.

b) Italicize the titles of books, essays, pamphlets, published documents, newspapers, periodicals, but not the titles of articles, which should be placed in double quotation marks.

c) Boldface type may be used to highlight important terms when they are first introduced and defined.

d) Italics, boldface type and upper-case letters (full capitals) should not be used to emphasize words or sentences. For boldface in particular see item c).

e) In some exceptional cases, underlining may be required in the final printed text. For example, underlining is acceptable in textual analysis to show high pitch and/or volume. Please use underlining only in this case and not in order to designate italics.

f) Small capitals are used for two purposes:

- for elements of semantic representations, acronyms and other words that appear in small capitals at every occurrence;
- to designate phonological/phonetic stress.

8.7. Punctuation

a) Single quotation marks are used only for the translation of words or phrases from languages other than English (for example, *cogito* `I think').

b) Use double quotation marks for direct quotations.

c) Use double quotation marks for "qualified" words or phrases.

d) Quotation marks should be placed inside punctuation when a word or part of a sentence is quoted, or when the title of an article, a contribution to a book, a poem, etc., is quoted. They are placed outside punctuation when complete sentences are cited.

e) Words containing prefixes are written solid, without hyphens, when no misreading will result: "antimentalism", "subdialect". The prefix is followed by a hyphen when the next element begins with a capital letter: "proto-Germanic".

f) Use hyphens only in words that require a hyphen no matter where they appear on the line, and place no spaces before or after the hyphen: "a deep-green sea".

In the printed material a dash that is longer than a hyphen ("en" dash) will be used between continuing or inclusive dates and numbers: "1965-1966", pages "5-8". However, since not all software has an "en" dash please do not use it even if it is available to you. Instead, use a single hyphen with one space on either side of it: "1965 - 1966".

Similarly, the longer parenthetical dash ("em" dash) should be represented by a double hyphen with no space before or after it--as shown here.

g) When referring to a book with more than one author, use the "&", as in "Dale & Werner" or "Smith, Brown & Jones". In the case of more than three authors use "et al.": "Smith et al.".

8.8. Abbreviations

a) Avoid abbreviations; they often pose severe problems to readers not completely familiar with the language of a text. Please limit your use of abbreviations to the few extremely common ones, such as "i.e., e.g., et al., etc.".

b) In general, language names must not be abbreviated except when prefixed to linguistic forms cited; thus "the meaning of OEngl. *guma*" is acceptable but "the meaning of *guma* in OEngl." is not. The latter must be rendered as "the meaning of guma in Old English".

c) Abbreviations ending in a small letter have a period following them (OFr., Gk., Lat.); those ending in a capital letter do not (MHG, OCS, OE).

d) If more than one abbreviation is acceptable, select one and use it consistently throughout the text.

8.9. Examples and foreign words

a) A letter, word, phrase, or sentence cited as a linguistic example or as the subject of discussion appears in italics; do not use quotation marks for this purpose.

b) Cited forms in a foreign language should be followed at least at first occurrence by a gloss in single quotation marks. No comma follows the gloss unless it is required by the sentence as a whole, e.g., "Lat. *ovis* `sheep', *equus* `horse', and *canis* `dog' are nouns." (Note that the commas follow the closing quotation mark.)

c) Special (e.g., phonetic or phonemic) symbols and other characters which cannot be produced by your software should be drawn in by hand on the hard copy. They need to be coded generically on the disk, and you must provide a list of the codes with the symbols they represent clearly drawn and identified by name.

d) Displayed examples should begin flush left. They should be separated from the preceding and the following text by one line of space and numbered consecutively throughout an article or, in the case of a monograph, throughout the whole text. Place the number in brackets, but not the letter following it. A period is used after the letter and at the end of an example, if it is a sentence.

(1) a. I sent the artifacts to an anthropologist.
b. I sent to an anthropologist the artifacts that had been in the attic.

(2) ??I sent to an anthropologist the artifacts.

Examples from languages other than English must have interlinear glosses below them and, in addition, a full free translation:

Original language in plain script
Gloss in smaller typeface
`Translation in single quotation marks'

(3) mampianatra angilisy an-d Rabe aho
Cause-learn English ACC-Rabe I
'I am teaching Rabe English.'

Type the glosses in exact alignment with the words in the example, just as they are to appear in the final version. Be sure to use tabs to align the examples and glosses. Do not use the space bar.

e) Take care not to overrun the line, and indicate a suitable line-break where necessary, either through a code on the disk or by marking the printout.

f) References to examples in the text should take the form "see, for example, (1a) and (1b)", with both number and letter in parentheses.

8.10. Transliteration

Examples from languages that do not use the Latin script should be presented in an accepted transliteration. In case of doubt, the editors should be consulted in advance.

8.11. Notes

a) Format notes using your program, with note numbers as superscripts in your text. Use the program to number notes automatically.

b) Note numbers in the text are indicated by a raised superscript devoid of punctuation or parentheses.

c) All punctuation marks, including closing parentheses, precede note numbers in the text.

d) In a work by a single author, notes are numbered serially throughout the text and should be placed in a separate section at the end of the text, before the references section.

e) For a work consisting of articles by several authors, notes are numbered serially throughout each article and should be placed at the end of each article, before the references.

8.12. Notes sheet

Numbers should appear left, followed by a period; the text is indented:

1. Text
   Text
2. Text
   Text

Be sure to use the tab or hanging indent to align the notes, not the space bar.

8.13. Citations

a) Full references for the literature cited are given in the references at the end of the manuscript.

b) In the text itself, only brief citations are included. These take the form "Hockett 1964: 240 - 241". Note that the page number or numbers given are those of the passage in an article or book to which reference is actually made, not to the entire work. Avoid global references such as "Chomsky 1965".

c) When reference is made to inclusive page numbers no digits are dropped, i.e.: "240 - 241", not "240 - 41" or "240 - 1". Do not use "f." or "ff." to indicate page numbers. The exact page numbers must be given in full.

d) Citations of books by more than one author take the form "Bartsch & Vennemann 1982: 1", "Smith, Brown & Jones 1989: 2". The names are separated by "&". For books with more than three authors "et al." is used in the text; the names of all the authors are given in the references.

e) When listing several authors, separate them with semi-colons; commas separate the dates of works by one author, as in "Smith 1984: 56, 1985a: 25; Brown 1977: 249; Jones 1990: 332".

f) When a citation refers to a work consisting of more than one volume, the form "1976, 1: 210" is used.

g) Reprint editions are cited as follows: "Gabelentz [1972]: 70" or, if it is important that the original date of publication be included in the text, "1901 [1972]: 70".

h) For brief citations, use initials or first names only when it is necessary to distinguish two or more authors with identical surnames.

i) If an author's name is part of the running text, use the form: "Bloomfield (1933: 264) introduced the term ...."

j) Do not use "ibid." and "op. cit.". Instead, repeat the brief citation.

8.14. Tables, figures, and illustrations

Submit all tables, figures, and captions in computer files that are separate from the main text, with notes indicating where each table or figure is to be placed. For example, in the text, "Table 1 about here", and in the table file, "Table 1 in section 1.1.3". As a rule, the typesetters will set your tables conventionally and insert them in the text, because conversion of tables is technically not usually feasible.

8.14.1. Tables

a) Leave ample space between columns, and double space all entries. Please do not use vertical lines.

b) Column headings should be short, so as to stand clearly above the columns. If you need longer headings, represent them by roman numbers and explain these in the text

pertaining to the table.

c) If two or more tables appear, number them and refer to them by number. Do not speak of the "preceding" or the "following" table, or "Table 3 below", since positioning of the tables may be affected by the page breaks in the final printed version. This also applies to figures and illustrations.

d) Each table should have a legend above it. The legend should contain the table number and a concise title in the form:

Table 1. Caption in roman type

If a (brief) explanation or comment is required, give it under the table.

e) Notes in tables are indicated by raised lower-case letters or numbers in superscript, the notes appearing below the table.

8.14.2. Figures

a) All figures must be provided in reproducible form. If your figures are not computer-generated, draw them with India ink on tracing paper; only one figure should appear on each sheet. All figures should be drawn to the same scale in such a way that the reduction (if necessary) can be the same for all drawings.

b) All figures should be numbered consecutively with arabic numerals:

Figure 1. Caption in roman type

Use simply "Figure 1, Figure 2, Figure 3", rather than a numbering system relating to the subsections of your work, such as "Figure 1.1", etc.

c) Type the captions underneath the figures; captions for all figures should also be listed on a separate sheet of paper.

8.14.3. Illustrations

If photographs have to be inserted, the print and the negative (or microfilm) should be provided. Do not send photocopies.

8.15. Orthography

Both American English and British English forms are acceptable, but spelling must be consistent throughout. In the case of manuscripts consisting of contributions by a number of authors, the volume editor must decide in favor of American or British English and edit the entire manuscript accordingly.

8.16. Obtaining permissions

It is the author's responsibility to request any permission required for the use of material owned by others. When all permissions have been received, the author should send them, or copies of them, to the publisher, who will note and comply with any special provisions regarding credit lines contained in them.

Sample references

Anttila, Raimo (1972). *An introduction to historical and comparative linguistics*. New York: Macmillan.

Bartsch, Renate & Theo Vennemann (1982). *Grundzüge der Sprachtheorie: Eine linguistische Einführung*. Tübingen: Niemeyer.

Eaton, Roger, Olga Fischer, Willem Koopman & Frederike van der Leek (eds.) (1985). *Papers from the fourth international conference on English Historical Linguistics*. (Current Issues in Linguistic Theory 41.) Amsterdam: Benjamins.

Fisiak, Jacek (1980). "Some notes concerning contrastive linguistics", *Association Internationale de Linguistique Appliquée Bulletin* 27: 1 - 17.

Fisiak, Jacek (1983). "Present trends in contrastive linguistics", in: Kari Sajavaara (ed.), 9 - 38.

Fisiak, Jacek (ed.) (1984). *Historical syntax*. (Trends in Linguistics: Studies and Monographs 23.) Berlin: Mouton de Gruyter.

Gabelentz, Georg von der (1901). *Die Sprachwissenschaft: Ihre Aufgaben, Methoden und bisherigen Ergebnisse*. (2nd edition.) Leipzig: Tauchnitz. Reprinted Tübingen: Narr (1972).

Goddard, Ives (1975). "Algonquian, Wiyot, and Yurok: Proving a distant genetic relationship", in: M. Dale Kinkade & Oswald Werner (eds.), 249 - 262.

Golla, Victor (1987). Review of Greenberg 1987a. *Current Anthropology* 28: 657 - 659.

Greenberg, Joseph H. (1987a). *Language in the Americas*. Stanford: Stanford University Press.

Greenberg, Joseph H. (1987b). Reply. *Current Anthropology* 28: 664 - 666.

Greenberg, Joseph H. (ed.) (1978). *Universals of language*. Cambridge, Mass.: The MIT Press.

Hashimoto, Mantaro (1987). "Kokusaigo toshite no kango to kanji" [Chinese characters and Chinese words as internationalisms], in: M. Hashimoto & T. Suzuki & H. Yamada, *Kanji minzoku no ketsudan* [Assessing the Kanji race]. Tokyo: Taishukan, 327 - 360.

Heath, Shirley Brice (in press). "The essay in English: Readers and writers in dialogue", in: M. Macovski (ed.), *Textual voices, vocative texts*. New York: Oxford University Press.

Hoenigswald, Henry M. (1978). "Are there universals of linguistic change?" in: Joseph H. Greenberg (ed.), 30 - 52.

Jespersen, Otto (1927). *A modern English grammar, Part III: Syntax*. London: Allen and Unwin.

Jones, Daniel (1950). *An English pronouncing dictionary*. (11th edition.) London: Dent.

Kinkade, M. Dale & Oswald Werner (eds.) (1975). *Linguistics and anthropology: In honor of C.F. Voegelin*. Lisse: de Ridder.

Lunt, Horace G. (1952). *A grammar of the Macedonian literary language*. Skopje [No indication of publisher.]

Meier, Hans Heinrich (1967). "The lag of relative *who* in the nominative", *Neophilologus* 51: 277 - 286.

Meillet, Antoine (1926 & 1936). *Linguistique historique et linguistique générale*. Vols. 1 -2. Paris: Champion/Klincksieck.

Parret, Herman (ed.) (1976). *History of linguistic thought and contemporary linguistics*. Berlin & New York: Walter de Gruyter.

Pott, August Friedrich (1833 & 1836). *Etymologische Forschungen auf dem Gebiet der indogermanischen Sprachen*. 2 vols. Lemgo: Meyer.

Romaine, Suzanne (1984). "Towards a typology of relative clause formation strategies in Germanic", in: Jacek Fisiak (ed.), 437 - 470.

Sajavaara, Kari (ed.) (1983). *Cross-language analysis and second language acquisition* 1. (Jyväskylä Cross-Language Studies 9.) Jyväskylä: University of Jyväskylä.

Sajavaara, Kari (ed.) (forthcoming). "Psycholinguistic testing of transfer in foreignlanguage speech processing".

Sapir, Edward (1929). "Central and North American languages", *Encyclopaedia Britannica*. (14th edition.) London & New York: Encyclopaedia Britannica Company. 5: 138 - 141. Reprinted in: David G. Mandelbaum (ed.), *Selected writings of Edward Sapir*. Berkeley: University of California Press, 1951, 169 - 178.

Sapir, Edward (1937). "The contribution of psychiatry to an understanding of behavior in society", *American Journal of Sociology* 42: 862 - 870.

Senn, Alfred (1966). *Handbuch der litauischen Sprache* 1. Heidelberg: Winter.

Silver, Shirley (1966). The Shasta language. [Unpublished Ph.D. dissertation, University of California at Berkeley.]

A supplement to the Oxford English dictionary. Ed. R. W. Burchfield. Vols. 1 & 4. Oxford: Clarendon Press.

Talmy, Leonard (n.d.). A comparison of the order of morpheme-classes in the Atsugewi and the Kashaya verb. [Unpublished MS.]

Back to index

9. Bibliographical entries

Bibliographical entries may be stored in a database of the general structure indicated in section 4.1.1. A record is founded by a publication. The following six publication types are distinguished:

- independent: book
  - (1) monograph
  - collection of articles
    - (2) reader: responsible persons are editors
    - (3) essay collection (selected writings): responsible person is author
- dependent: article
  - (4) journal article
  - (5) article in collective work
- (6) unpublished (grey literature).

A collective work and an article contained therein constitute separate entries. The latter refers to the former.

9.1. General format

9.1.1. Structure of a bibliographical entry

The following is the maximum field structure of a record in a bibliographical database. The fields contain plain text. Typographical make-up, such as italics, quotation marks around titles, punctuation between pieces of information, are a matter of the style-sheet (see section 8) and are not entered into the database.

1. **Name 1**: Last name of first author/editor
2. **First name 1**: First name of first author/editor
3. **Name 2**: Last name of second author/editor (or `et al.')
4. **First name 2**: First name of second author/editor
5. **Publication type**: Abbreviation identifying type
6. **Year**: Year(s) of publication
7. **Title**: Main title of this entry
8. **Subtitle**: Subtitle and number of volumes
9. **Journal or reader**: For an article: name of journal or reference to collective work (in the same database)
10. **Volume and pages**: For an article: volume number of journal or collective work and page numbers occupied
11. **Place**: For a book: Town(s) of publisher
12. **Publisher**: For a book: Name of publishing company/-ies
13. **Series**: For a book: name of publisher's series, volume number
14. **Editions**: For a book: edition of this entry, earlier impressions; for an article: reference to reprints of this entry
15. **Original**: For revised editions, translations and reprints: reference to original edition
16. **Reviews**: Bibliographical data of reviews of this entry

17. **Area**: Geographical area to which the study is confined
18. **Languages**: Languages to which the study is devoted
19. **Descriptors**: Items from the terminological network (see section 2.2) which describe this entry
20. **Availability**: Owner, esp. library with shelf mark
21. **Comments**: Any comments, esp. summary of the entry
22. **Number**: Consecutive number according to entry time (for database administration)

9.1.2. Subset of fields in a record

The field structure of a given record is an appropriate subset of the above field structure. In the selection of this subset, the following considerations apply:

- #3 and 4 are used as they apply. In a database with flexible field structure, #1 and 2 could be repeated for any number of authors.
- #5 is needed to select the appropriate typographic style when printing a report.
- #9 and 10 are used for articles, #11 to 13 instead for independent publications; the two subsets are mutually exclusive.
- The information contained in #17 to 19 has a set structure. In a database with flexible field structure, the respective field can be repeated for each item. Otherwise, the elements of the set should be formally identified as such (e.g.: Languages: {Catalan} {Gallego}).

The following fields need to be filled in for each record, if the database is to work:

- #1 and 2 if there is an author
- #5, 6, 7
- either #9 and 10 or #11 and 12.

9.2. Example entries

In the following examples, fields are identified by the numbers used in section 9.1.1.

9.2.1. Monograph

1. Allen
2. W.Sidney
5. m
6. 1973
7. Accent and rhythm
8. Prosodic features of Latin and Greek: A study in theory and reconstruction
11. Cambridge
12. University Press
13. Cambridge Studies in Linguistics, 12
16. Newton 1975[A]
18. Latin
18. Ancient Greek
19. prosody
20. UB Bi: 15 NG 3 40.00 A 432
21. I. The general and theoretical background; II. The prosodies of Latin; III. The prosodies of Greek.

9.2.2. Essay collection

1. Benveniste
2. Emile
5. e
6. 1966
7. Problèmes de linguistique générale
11. Paris
12. Éd. Gallimard
13. Bibliothèque des Sciences Humaines
14. Engl.: Problems in general linguistics. Coral Gables, Fla.: Univ. of Miami Press, 1971. Germ.: Benveniste 1974[PI]
16. Mounin 1967[E]
19. European structuralism
20. UB Bi: 15 NA 101.00 B478

9.2.3. Reader

1. Davidson
2. Donald
3. Harman
4. Gilbert
5. eds.
6. 1972
7. Semantics of natural language
11. Dordrecht
12. D. Reidel
13. Synthese Library
15. Synthese 12: 249-487; 22: 1-289
16. Leist 1974
19. logical semantics

9.2.4. Article in collective work

1. McCawley
2. James D.
5. s
6. 1972
7. A program for logic
9. Davidson & Harman (eds.)
10. 498-544
19. semantic representation
19. natural logic

9.2.5. Journal article

1. Benveniste
2. Emile
5. j
6. 1949
7. Le système sublogique des prépositions en latin
9. Travaux du Cercle Linguistique de Copenhague

10. 5:177-184
14. Benveniste 1966: 132-139
18. Latin
19. local preposition

9.2.6. Unpublished work

1. Bakker
2. Dik
3. Siewierska
4. Anna
5. u
6. 1991
7. A database system for language typology
11. Strasbourg
12. Fondation Européenne de la Science
13. EUROTYP Working Papers, II, 3
19. typological methodology
19. database
19. word order typology
20. Ö.D.

Back to index