

In Praise of Simplicity:
Lexicographic Lightweight Markup Language

Vladimír Benko

vladimir.benko@juls.savba.sk

Slovak Academy of Sciences, Ľ. Štúr Institute of Linguistics



XVIII EURALEX International Congress

Ljubljana | 17–21 July 2018

Representation of Lexicographic Data

(1) “Easy way”: **Formatted Text**

- “WYSIWYG-only” environment (usually within a word processor such as Microsoft Word)
- Information fields indicated by typography (font types, sizes and colours), punctuation and formatting and/or styles

Advantages

- Necessary software is (usually) already available
- No extensive lexicographic team training is required
- Only limited (human) IT resource is needed
- No conversion to (paper form) is needed
- **Lexicographers’ mental model of a dictionary entry is close to what they can see at the screen**

Representation of Lexicographic Data

(1) “Easy way”: **Formatted Text**

- “WYSIWYG-only” environment (usually within a word processor such as Microsoft Word)
- Information fields indicated by typography (font types, sizes and colours), punctuation and formatting and/or styles

Disadvantages

- Complex structures cannot be encoded explicitly
- Structure uniformity is difficult to impose
- Data validation is difficult
- Data reusability is questionable

Representation of Lexicographic Data

(1) “Easy Way”: Formatted Text

БИ^СЕР, СУЩ; МУЖСК; -а.

ПРИМЕРЫ. *Бусы из бисера; жакет с отделкой из бисера; вышивать бисером.*

ЗНАЧЕНИЕ. ‘Маленькие, размером от зерна до мелкой ягоды, шарики, цилиндры или другой похожей формы объекты со сквозными отверстиями, которые пришиваются на ткань для ее украшения или нанизываются на что-л. и носятся как украшение’.

КОММЕНТАРИИ. Образные употребления применительно к каплям влаги: *сверкающий бисер росы; Все его лицо было покрыто мелким бисером пота* (А. Житков).

СОЧЕТАЕМОСТЬ. *Цветной <разноцветный, белый, красный, черный, розовый> бисер, прозрачный <непрозрачный> бисер; крупный <мелкий> бисер; стеклянный <жемчужный, серебряный, золотой, деревянный, янтарный, пластиковый> бисер; зерна бисера; браслеты <четки> из бисера; расшитый <вышитый, шитый, усыпанный> бисером; нанизывать <пришивать> бисер; украшать бисером; плести из бисера браслеты.*

АНАЛ: жемчуг; стеклярус; бусинки; ДЕР: бисерина, бисеринка; бисероплетение; бисерный.

ИЛЛЮСТРАЦИИ. *Рюкзак, фляга на поясе, на руке компас, индейский браслетик из бисера* (В. Песков, Б. Стрельников). *Она страдала мигренью из-за проклятого бисера: целыми днями плела украшения на продажу* (Д. Быков). *Все трое расселись вокруг большого стола, застеленного шитой бисером скатертью* (Л. Юзефович). *Взял ночные туфли, вышитые бисером и купленные еще в Нижнем Новгороде у рукодела-монаха Печорского монастыря* (А. Мариенгоф).

{ } *Метать бисер перед свиньями* библ. ‘говорить или объяснять что-то важное человеку, который не может или не хочет этого понять и оценить’: *Не мечите бисера перед свиньями; Тут не место стихам. Хватит метать бисер перед свиньями* (Ю. Нагибин). [В.А.]

Representation of Lexicographic Data

(1) “Easy Way”: Formatted Text

БИСЕР, СУЩ; МУЖСК; *-а*.

Бусы из бисера; жакет с отделкой из бисера; вышивать бисером.

ЗНАЧЕНИЕ. ‘Маленькие, размером от зерна до мелкой ягоды, шарики, цилиндры или другой похожей формы объекты со сквозными отверстиями, которые пришиваются на ткань для ее украшения или нанизываются на что-л. и носятся как украшение’.

👉 **Образные употребления применительно к каплям влаги:**
сверкающий бисер росы; Все его лицо было покрыто мелким бисером пота (А. Житков).

СОЧЕТАЕМОСТЬ. *Цветной <разноцветный, белый, красный, черный, розовый> бисер, прозрачный <непрозрачный> бисер; крупный <мелкий> бисер; стеклянный <жемчужный, серебряный, золотой, деревянный, янтарный, пластиковый> бисер; зерна бисера; браслеты <четки> из бисера; расшитый <вышитый, шитый, усыпанный> бисером; нанизывать <пришивать> бисер; украшать бисером; плести из бисера браслеты.*

АНА: *жемчуг; стеклярус; бусинки; ДЕР:* *бисерина, бисеринка; бисероплетение; бисерный.*

📖 *Рюкзак, фляга на поясе, на руке компас, индейский браслетик из бисера (В. Песков, Б. Стрельников). Она страдала мигренью из-за проклятого бисера: целыми днями плела украшения на продажу (Д. Быков). Все трое расселись вокруг*

Representation of Lexicographic Data

(2) “Difficult/Complex Way”: **Structured Text**

- Generalized markup language (SGML, XML, TEI-P5, ISO 1951:2007, ...) plus data definition tools (DTD, Schema, ...)
- Two working spaces: “structured” and “presentation”

Advantages

- Complex structures can be encoded explicitly
- Uniformity can be imposed
- Data validation possible
- Data reusability guaranteed

Representation of Lexicographic Data

(2) “Difficult/Complex Way”: **Structured Text**

- Generalized markup language (SGML, XML, TEI-P5, ISO 1951:2007, ...) plus data definition tools (DTD, Schema, ...)
- Two working spaces: “structured” and “presentation”

Disadvantages

- Specialized (often expensive) software needed
- Extensive training of the lexicographic team required
- Intensive (human) IT engagement necessary
- Data reusability guaranteed
- Data conversion always required
- **Lexicographers’ mental model of a dictionary entry differs significantly from the data that are working with**

Representation of Lexicographic Data

```
1 <div type="dictionary" xml:id="lachbaum" xml:lang="de" rend="unmarked">
2   <entry>
3     <form type="headword">
4       <gram type="art_m">Der </gram>
5       <orth>Lachbaum <c type="$,">,</c></orth>
6     </form>
7     <gramGrp>
8       <number value="sing"/>
9       <case value="gen">
10        <oVar type="gen"><oRef/>-es </oVar>
11        <c type="$,">,</c>
12      </case>
13      <number value="plur">Mz.</number>
14      <case value="nom">
15        <oVar type="pl"><oRef/>-bäume </oVar>
16        <c type="$,">,</c>
17      </case>
18    </gramGrp>
19    <usg type="dom">im Forstwesen <c type="$.">.</c></usg>
20    <sense>
21      <sense n="1" level="1" rend="numbered">
22        <c type="ordnungselement ">1) </c>
23        <def>Ein an sei-<lb/>
24          ner Borke aufgehauener Nadelholzbaum<c type="$,">,</c> aus welchem man Harz <lb/>
25          rinnen läßt<c type="$.">;</c>
26        </def>
27      </sense>
28      <sense n="2" level="1" rend="numbered">
29        <c type="ordnungselement">2) </c>
30        <def>ein Baum<c type="$,">,</c> welchen man auf der Grenze des Wal-<lb/>
31          des neben den Grenzsteinen stehen läßt und auf zwei Seiten mit drei <lb/>
32          Hieben bezeichnet<c type="$,">,</c> um den Grenzstein desto besser finden
33          zu können<c type="$.">;</c><lb/>
34        </def>
```


Representation of Lexicographic Data

```
1 -<div type="dictionary" xml:id="lachbaum" xml:lang="de" rend="unmarked">
2 -  <entry>
3 -    <form type="headword">
4 -      <gram type="art_m">Der </gram>
5 -      <orth>Lachbaum <c type="$,">,</c></orth>
6 -    </form>
7 -    <gramGrp>
8 -      <number value="sing"/>
9 -      <case value="gen">
10 -        <oVar type="gen"><oRef/>-es </oVar>
11 -        <c type="$,">,</c>
12 -      </case>
13 -      <number value="plur">Mz.</number>
14 -      <case value="nom">
15 -        <oVar type="pl"><oRef/>-bäume </oVar>
16 -        <c type="$,">,</c>
17 -      </case>
18 -    </gramGrp>
19 -    <usg type="dom">im Forstwesen <c type="$.">.</c></usg>
20 -    <sense>
21 -      <sense n="1" level="1" rend="numbered">
22 -        <c type="ordnungselement ">1) </c>
23 -        <def>Ein an sei-<lb/>
24 -          ner Borke aufgehauener Nadelholzbaum<c type="$,">,</c> aus welchem man Harz <lb/>
25 -          rinnen läßt<c type="$.">;</c>
26 -        </def>
27 -      </sense>
28 -      <sense n="2" level="1" rend="numbered">
29 -        <c type="ordnungselement">2) </c>
30 -        <def>ein Baum<c type="$,">,</c> welchen man auf der Grenze des Wal-<lb/>
31 -          des neben den Grenzsteinen stehen läßt und auf zwei Seiten mit drei <lb/>
32 -          Hieben bezeichnet<c type="$,">,</c> um den Grenzstein desto besser finden
33 -          zu können<c type="$.">;</c></def>
```

Representation of Lexicographic Data

Why (even today) some opt for “easy way”?

“Difficult way” is

- Too complex/complicated
- Too expensive (software, training, IT support)
- Older team members are reluctant to learn something new

For long-term projects

- *“Cannot afford changing the workflow each decade”*

Representation of Lexicographic Data

Possible compromise

Lightweight Markup Language

A *lightweight markup language (LML)*, also termed a **simple** or **humane markup language**, is a **markup language** with simple, unobtrusive **syntax**.

*It is designed to be easy to write using any generic **text editor** and easy to read in its raw form.*

Lightweight markup languages are used in applications where it may be necessary to read the raw document as well as the final rendered output.

[en.wikipedia.org, 2018]

Representation of Lexicographic Data

Possible compromise

Lightweight Markup Language



The screenshot shows a Wikipedia article page. At the top left is the Wikipedia logo, a globe with puzzle pieces, and the text 'WIKIPEDIA The Free Encyclopedia'. Below the logo is a sidebar with navigation links: Main page, Contents, Featured content, Current events, Random article, Donate to Wikipedia, Wikipedia store, Interaction, and Help. The main content area has a title 'Editing Lightweight markup language' and a warning: 'Content that violates any copyrights will be deleted. Encyclopedic content must be verifiable. Work submitted to Wikipedia can be edited, used, and redistributed—by anyone—subject to certain terms and conditions.' Below the warning is a toolbar with icons for bold, italic, link, image, and other editing functions. The article text is in a monospaced font and contains a code block for a template and a paragraph of text.

Grano 2 3 Talk Sandbox Preferences Beta Watchlist Contributions Log out

Article Talk Read Edit source View history

Editing Lightweight markup language

Content that violates any copyrights will be deleted. Encyclopedic content must be verifiable. Work submitted to Wikipedia can be edited, used, and redistributed—by anyone—subject to certain terms and conditions.

B *I* [Advanced](#) [Special characters](#) [Help](#) [Cite](#)

```
{{primary sources|date=December 2007}}
```

A '''lightweight markup language''' ('''LML'''), also termed a '''simple''' or '''humane markup language''', is a [[markup language]] with simple, unobtrusive [[syntax (programming languages)|syntax]]. It is designed to be easy to write using any generic [[text editor]] and easy to read in its raw form. Lightweight markup languages are used in applications where it may be necessary to read the raw document as well as the final rendered output.

For instance, a person downloading a software library might prefer to read the documentation in a text editor rather than a web browser. Another application for such languages is to provide for data entry in [[web publishing|web-based publishing]], such as [[weblog]]s and [[wiki]]s, where the input interface is a simple [[text box]]. The server software then converts the input into a common [[document markup language]] like [[HTML]].

Lexicographic Lightweight Markup Language

“Historical background”

- Developed in early 1990s to represent a Slovak monolingual dictionary converted from typesetting format (Ferranti)
- Requirements:
 - MS-DOS only environment
 - no network access
 - simple (command line) validation tools
- Later used in preparation of new editions of that dictionary
- In late 1990s applied to all lexicographic projects at our Institute as a “temporary solution” that for various reasons survived until today ;-)

Lexicographic Lightweight Markup Language

“Historical background”

Dictionaries using *LLML* (Ľ. Štúr Institute of Linguistics)

- **Concise Dictionary of the Slovak Language** (3rd and 4th edition)
- **Slovak Synonym Dictionary** (3 editions)
- **Dictionary of Slovak Dialects** (2 volumes, the 3rd in preparation)
- **Historical dictionary of the Slovak Language** (3rd to 7th volume, plus Appendix)
- **Dictionary of the Contemporary Slovak Language** (3 volumes, the 4th in preparation)

Lexicographic Lightweight Markup Language

- “Historical background”

Dictionaries using *LLML* (Slovak Educational Publishers)

- **Dictionary of Loanwords** (2 editions)
- **Passport English-Slovak Dictionary** (licensed by K Dictionaries)
- **Compact English-Slovak & Slovak English Dictionary** (2 editions)
- **Russian-Slovak Dictionary of Idioms**

Lexicographic Lightweight Markup Language

“Historical background”

Other use of *LLML*

- **Encyclopaedia of the Slovak Folk Culture** (Institute of Ethnology)
- **Global Password Series** (K Dictionaries, initial 30 language conversion from typographical format)

Lexicographic Lightweight Markup Language

<i>LLML Element</i>	<i>Default rendering</i>
"headword"	headword
"headword^1"	headword ¹ (headword with index)
"%substandard headword"	substandard headword
"*incorrect headword"	*incorrect headword
"~crossref headword"	crossref headword
[pronunciation]	[pronunciation]
*PoS label	pos label
other label	other label
<etymology>	<etymology>
'example text'	<i>example text</i>
[*reference]	[reference]
{1}, {2}, ...	1. , 2. , ... (sense numbers)
{M}, {T}, ...	□, □, ... (special symbols indicating “structural breaks” in entry structure)
(unmarked)	(unmarked) ... definitions, explanations, etc.

Lexicographic Lightweight Markup Language

309 !10260
310 "légia" -ie |pl. G| -ií |D| -iám |L| -iách |*ž.| <lat.>
311 {1} |hist., voj.| najväčšia a hlavná bojová jednotka
312 rímskej armády, ktorú tvorilo niekoľko kohort: 'rímske
313 légie'; 'l. cisára Marca Aurelia'; 'veliteľ légiám'; 'poraziť
314 légie'; 'privolať na pomoc légie z Východu'; 'Rím,
315 zákonodarca a vládca sveta, sídlo nepremožiteľných
316 légií.' [*Anton Hlinka]
317 {2} |voj.| dobrovoľná vojenská jednotka: 'veliteľ,
318 príslušník légie'; 'v roku 1916 vstúpil do
319 československých légií v Rusku'
320 {T} |hist.| 'Biela légia' protikomunistické ilegálne
321 hnutie pôsobiace na Slovensku v r. 1948 - 1955; 'légia
322 Kondor' nacistická vojenská jednotka vyslaná Hitlerom
323 do Španielska počas španielskej občianskej vojny
324 {M} 'Cudzinecká légia' francúzska špeciálna jednotka
325 tvorená zahraničnými dobrovoľníkmi, súčasť francúzskej
326 armády, v súčasnosti nasadzovaná v rámci mierových
327 humanitných operácií vojenských síl OSN
328 {3} |expr.| veľký počet niečoho, množstvo ľudí, dav:
329 'l. básnikov'; 'Mačiek sa vrátila celá légia, lebo sa
330 niekde nakotili a spätnásobili.' [*Š. Žáry]; 'Federálne
331 a miestne vlády zamestnávajú celé légie inšpektorov,
332 ktorí udeľujú vysoké pokuty.' [*HN 2003]
333 ?a IKPMV
334 ?b NK
335 ?? Anton Hlinka - Ozvena slova 1 - Blahozvešť v horizonte
336 ?? Ľudskej skúsenosti 2, 1996
337 ?? Štefan Žáry - Apeninský vzduch, 1984
338 ?? Teofil Klas - Putovanie do Loreta, 1999

Lexicographic Lightweight Markup Language

```
309 !10260
310 "légia" -ie |pl. G| -ií |D| -iám |L| -iách |*ž. | <lat.>
311 {1} |hist., voj. | najväčšia a hlavná bojová jednotka
312 rímskej armády, ktorú tvorilo niekoľko kohort: 'rímske
313 légie'; 'l. cisára Marca Aurelia'; 'veliteľ légiám'; 'poraziť
314 légie'; 'privolať na pomoc légie z Východu'; 'Rím,
315 zákonodarca a vládca sveta, sídlo nepremožiteľných
316 légií.' [*Anton Hlinka]
317 {2} |voj. | dobrovoľná vojenská jednotka: 'veliteľ,
318 príslušník légie'; 'v roku 1916 vstúpil do
319 československých légií v Rusku'
320 {T} |hist. | 'Biela légia' protikomunistické ilegálne
321 hnutie pôsobiace na Slovensku v r. 1948 - 1955; 'légia
322 Kondor' nacistická vojenská jednotka vyslaná Hitlerom
323 do Španielska počas španielskej občianskej vojny
324 {M} 'Cudzinecká légia' francúzska špeciálna jednotka
325 tvorená zahraničnými dobrovoľníkmi, súčasť francúzskej
326 armády, v súčasnosti nasadzovaná v rámci mierových
327 humanitných operácií vojenských síl OSN
328 {3} |expr. | veľký počet niečoho, množstvo ľudí, dav:
329 'l. básnikov'; 'Mačiek sa vrátila celá légia, lebo sa
330 niekde nakotili a spätnásobili.' [*Š. Žáry]; 'Federálne
331 a miestne vlády zamestnávajú celé légie inšpektorov,
332 ktorí udeľujú vysoké pokuty.' [*HN 2003]
333 ?a IKPMV
334 ?b NK
335 ?? Anton Hlinka - Ozvena slova 1 - Blahozvešť v horizonte
336 ?? Ľudskej skúsenosti 2, 1996
337 ?? Štefan Žáry - Apeninský vzduch, 1984
338 ?? Teofil Klas - Putovanie do Loreta, 1999
```

Lexicographic Lightweight Markup Language

Data validation (I)

- **Well-formedness, elementary syntax a typography**

Command-line tool to be used by lexicographers themselves

```
157 ***** po čiarke má byť medzera
    157:  {1} |lit.| literárny žáner o živote svätých, v ktorom

228-231 ***** číslo významu mimo poradia
    228:  {1} vlastnosť toho, čo sa zakladá na legendách,
    229:  vymyslených, neskutočných veciach: '1. príbehu';
    230:  'literatúra s črtami legendárnosti'
    231:  {3} vlastnosť toho, čo je nezabudnuteľné, výnimočné,
```

Lexicographic Lightweight Markup Language

Data validation (II)

- **“Batch mode”**

4	m00_majka_0612ao_w5_003360_majstrovat_____	"[*M. Beniak]"
1	n40_nekonv0802hs_w4_040000_nekonvenčne_____	"[*M. Burlas]"
17	m20_mladic0912as_w2_029370_mladučký_____	"[*M. Bútorá]"
2	m30_monofr0803dz_w5_037510_monolit_____	"[*M. Bútorá - M. Ivantyšyn]"
1	n00_nadsen0912jk_w3_007840_nadúvať_sa_____	"[*M. Čeretková - Gállová]"
2	o20_odober0904jh_w3_026710_odopierať^1_____	"[*M. Čeretková-Gallová]"
26	m00_masten0903ao_w3_011460_maškrtnica_____	"[*M. Čeretková-Gállová]"
5	n00_naadso0612jk_w5_001400_nábeh_____	"[*M. Čeretková-Gálová]"
7	m00_mahzo_0612ao_w5_002780_majestátny_____	"[*M. Danieliszová-Jesenická]"
1	m40_mucnat0811nj_w5_045640_%mukel_____	"[*M. Dudinský]"
2	n30_navabi0802ep_w4_030390_návej_____	"[*M. Dudok]"
30	m00_majka_0612ao_w5_003520_makarón_____	"[*M. Ďuríčková]"
6	n00_naczuc0612jk_w3_003800_načúvať_____	"[*M. Dziaková-Böhmová]"
6	m20_miagac0806as_w5_022670_miašť_____	"[*M. Dzvóník]"
1	n00_nahlad1108jk_w1_009930_naho_____	"[*M. Fazekašová]"
1	o50_oplies1102jh_w1_052920_opovažovať_sa_____	"[*M. Feduš]"
35	m20_mitoch0910as_w1_028481_mitrha_____	"[*M. Ferko]"
28	m20_mitoch0910as_w1_028760_mizéria_____	"[*M. Figuli]"
1	n40_neobme0901hs_w4_043640_neodvratne_____	"[*M. Gavenda]"

Lexicographic Lightweight Markup Language

Data validation (II)

- **“Batch mode”**

4	m00_majka_0612ao_w5_003360_majstrovat_____	"[*M. Beniak]"
1	n40_nekonv0802hs_w4_040000_nekonvenčne_____	"[*M. Burlas]"
17	m20_mladic0912as_w2_029370_mladučky_____	"[*M. Bútorá]"
2	m30_monofr0803dz_w5_037510_monolit_____	"[*M. Bútorá - M. Ivantyšyn]"
1	n00_nadsen0912jk_w3_007840_nadúvať_sa_____	"[*M. Čeretková - Gáľlová]"
2	o20_odober0904jh_w3_026710_odopierať^1_____	"[*M. Čeretková-Gáľlová]"
26	m00_masten0903ao_w3_011460_maškrtnica_____	"[*M. Čeretková-Gáľlová]"
5	n00_naadso0612jk_w5_001400_nábeh_____	"[*M. Čeretková-Gáľlová]"
7	m00_mahzo_0612ao_w5_002780_majestátny_____	"[*M. Danieliszová-Jesenická]"
1	m40_mucnat0811nj_w5_045640_%mukel_____	"[*M. Dudinský]"
2	n30_navabi0802ep_w4_030390_návej_____	"[*M. Dudok]"
30	m00_majka_0612ao_w5_003520_makarón_____	"[*M. Ďuríčková]"
6	n00_naczuc0612jk_w3_003800_načúvať_____	"[*M. Dziaková-Böhmová]"
6	m20_miagac0806as_w5_022670_miašť_____	"[*M. Dzvónik]"
1	n00_nahlad1108jk_w1_009930_naho_____	"[*M. Fazekašová]"
1	o50_oplies1102jh_w1_052920_opovažovať_sa_____	"[*M. Feduš]"
35	m20_mitoch0910as_w1_028481_mitrha_____	"[*M. Ferko]"
28	m20_mitoch0910as_w1_028760_mizéria_____	"[*M. Figuli]"
1	n40_neobme0901hs_w4_043640_neodvratne_____	"[*M. Gavenda]"

Lexicographic Lightweight Markup Language

Proofreading

310 10260 **légia** -ie pl. G -íí D -iám L -iách ž. <lat.>
311 {1} hist., voj. najväčšia a hlavná bojová jednotka
312 rímskej armády, ktorú tvorilo niekoľko kohort: *rímske*
313 *légie*; *L. cisára Marca Aurelia*; *veliteľ légiám*; *poraziť*
314 *légie*; *privolať na pomoc légie z Východu*; *Rím,*
315 *zákonodarca a vládca sveta, sídlo nepremožiteľných*
316 *légii*. [Anton Hlinka]
317 {2} voj. dobrovoľná vojenská jednotka: *veliteľ,*
318 *príslušník légie*; *v roku 1916 vstúpil do*
319 *československých légii v Rusku*
320 {T} hist. *Biela légia* protikomunistické ilegálne
321 hnutie pôsobiace na Slovensku v r. 1948 – 1955; *légia*
322 *Kondor* nacistická vojenská jednotka vyslaná Hitlerom
323 do Španielska počas španielskej občianskej vojny
324 {M} *Cudzinecká légia* francúzska špeciálna jednotka
325 tvorená zahraničnými dobrovoľníkmi, súčasť francúzskej
326 armády, v súčasnosti nasadzovaná v rámci mierových
327 humanitných operácií vojenských síl OSN
328 {3} expr. veľký počet niečoho, množstvo ľudí, dav:
329 *L. básnikov*; *Mačiek sa vrátila celá légia, lebo sa*
330 *niekde nakotili a späťnásobili*. [š. Žáry]; *Federálne*
331 *a miestne vlády zamestnávajú celé légie inšpektorov,*
332 *ktorí udeľujú vysoké pokuty*. [HN 2003]

333 ?a IKPMV

334 ?b NK

335 ?? Anton Hlinka — Ozvena slova 1 — Blahozvešť v horizonte

Lexicographic Lightweight Markup Language

Final conversion

légia *-ie* pl. G *-íí* D *-iám* L *-iách* ž. ⟨lat.⟩ **1.** hist., voj. ▶ najväčšia a hlavná bojová jednotka rímskej armády, ktorú tvorilo niekoľko kohort: *rímske légie; l. cisára Marca Aurelia; veliť légiám; poraziť légie; privoľat' na pomoc légie z Východu; Rím, zákonodarca a vládca sveta, sídlo nepremožiteľných légií.* [Anton Hlinka] **2.** voj. ▶ dobrovoľná vojenská jednotka: *veliteľ, príslušník légie; v roku 1916 vstúpil do československých légií v Rusku* □ hist. *Biela légia* protikomunistické ilegálne hnutie pôsobiace na Slovensku v r. 1948–1955; *légia Kondor* nacistická vojenská jednotka vyslaná Hitlerom do Španielska počas španielskej občianskej vojny □ *Cudzinecká légia* francúzska špeciálna jednotka tvorená zahraničnými dobrovoľníkmi, súčasť francúzskej armády, v súčasnosti nasadzovaná v rámci mierových humanitných operácií vojenských síl OSN **3.** expr. ▶ veľký počet niečoho, množstvo ľudí, dav: *l. básnikov; Mačiek sa vrátila celá légia, lebo sa niekde nakotili a späťnásobili.* [Š. Žáry]; *Federálne a miestne vlády zamestnávajú celé légie inšpektorov, ktorí udeľujú vysoké pokuty.* [HN 2003]

Lexicographic Lightweight Markup Language

Why **we** use *LLML* today

- Reluctant to change (for long-term projects)
- **In special cases: Retrodigitization**

1st and 2nd volume of the Historical Dictionary of the Slovak Language

- Scanning + OCR
- Output in RTF format with page layouts preserved
- Conversion into *LLML*
- Mass correction of metalanguage
- Spell-checking of definitions
- Manual corrections of texts in historical orthography

Lexicographic Lightweight Markup Language

Why we use *LLML* today

2919 "légia" |ž lat voj| 'vojenský útvar': Muhamed do Uher
2920 se nawratil, wogsko swé k lepssjmu pořádku při-
2921 westi vmjnil, vstanowil legii od černých pancýřů,
2922 černau rečenau [*StN 1786]
2923
2924 "legión" |m lat| 'římský pluk počítající 3-6-tisíc voja-
2925 kov légia': Cegsar pribeh s gednim legionem wog-
2926 ska [*NHí 1791]
2927
2928 "legislátor" |m lat práv| 'tvorca zákonov, zákonodarca':
2929 legifer: wydáwač práwa, ustanowitel práwa, wyda-
2930 watel zákona, legislátor [*KS 1763]
2931
2932 "legitimácia" |ž lat| 'písomné preukázanie totožnosti
2933 al. oprávnenia na niečo': aby jistota mohla se miti,
2934 na kterých mistech človek naleza se, jeden každý
2935 posel i o vyslani i o prichodu sveho čase legitimaci
2936 sobe má zaopatriti [*VIEDEŇ 1782 CM]; exactor takowé-
2937 to náklady bez důstatecneg legitimacie a převká-
2938 zanj přigaty nebude mócy [*TURIEC 1796]