

Novočeský lexikální archiv v průběhu let (1911–2011)

Hana Goláňová, Ústav pro jazyk český AV ČR, v. v. i., oddělení současné lexikologie a lexikografie

V letošním roce slavíme nejen 100. výročí založení Kanceláře Slovníku jazyka českého, ale zároveň i výročí vzniku novočeského lexikálního archivu (NLA), na jehož základě byl první výkladový „Slovník jazyka českého“ vytvořen. Excerptní listky dokumentují vývoj češtiny, zachycené zejména z psaných zdrojů, od počátku národního obrození do 90. let 20. století. Lístková část NLA představuje jedinečné kulturní dědictví, neboť obsahuje informace nedocenitelné hodnoty. Zde vás stručně seznámíme s materiálovými sbírkami NLA.

NLA v současné době obsahuje dvě samostatné části: starší **lístkový lexikální archiv** a nově budovaný **elektronický lexikální archiv**.

1. Lístkový lexikální archiv

Lístková část NLA je sestavena z materiálových vrstev, které vznikaly jednak pro potřeby tvorby jednotlivých slovníkových děl, jako synchronní excerptce z určitého období a také z praktických důvodů.

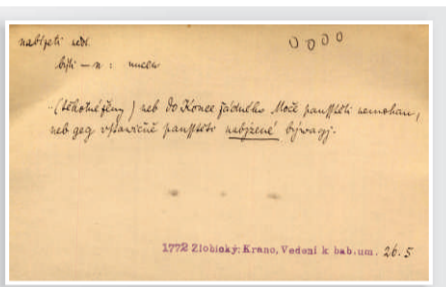


Základní, tzv. 1., materiál NLA sloužil především jako východisko pro zpracování „Slovníku jazyka českého“, později nazvaného *Příruční slovník jazyka českého* (PSJČ, 1935–1957), ale čerpa z něj i další slovníková díla.

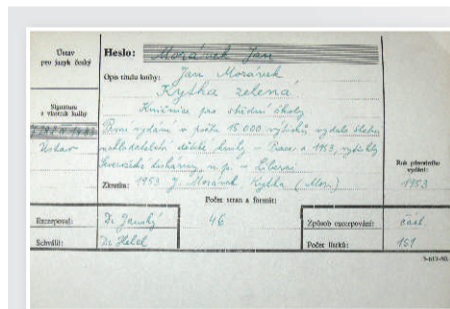
Lexikografická a dialektologická komise III. třídy České akademie věd a umění sestavila pro excerptory pravidla (*Návod pro sběratele materiálu k „Slovníku jazyka českého“*, 1911; *Dodatky a opravy*, 1913; *Pravidla pro sběratele látky k „Slovníku jazyka českého“*, 1917), na jejichž základě byla excerptce prováděna. Sběr lexikálního materiálu probíhal prostřednictvím lístkových výpisků zejména z umělecké literatury, mimo to byly výběrově

excerptovány noviny, časopisy a překladová literatura. Značná pozornost se věnovala i českému odbornému názvosloví (*Pravidla o vypisování odborných terminů pro Slovník jazyka českého*, 1917). Pro excerptci zdrojů byl jako dolní hranice stanoven rok 1770.

Vypisování materiálů určených k excerptci probíhalo trojím způsobem: podle excerptce úplné, částečné a tzv. paběrkování. **Úplná (totální) excerptce** spočívala v tom, že se jedenkrát vypsalo z autorova díla každé slovo (i slovo běžné) ve všech významech a významových odstínech. Později se častěji pracovalo metodou **částečné excerptce**, při níž byla pominuta běžná slovní zásoba a vypisovala se řidší slova, významy a vazby. Z toho postupu se vyvinula **excerptce speciálně zaměřená**. Pokud jde o **paběrkování**, vybíraly se pouze jazykové zvláštnosti.



Excerptní listek obvykle obsahuje tyto údaje: heslo s údajem slovního druhu, příp. podheslí, tj. vazbu nebo spojení, v němž se heslové slovo v citátu vyskytuje, citát s podtrženým heslovým slovem a bibliografický údaj.



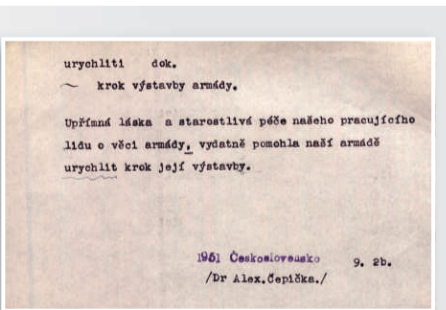
Součástí lístkového NLA jsou i tzv. zrcadla – záznamy obsahující bibliografický údaj o excerptovaných knihách a dalších materiálových zdrojích.

Když se v roce 1935 začalo s vydáváním PSJČ, měl NLA reprezentativní soubor pěti a půl milionů excerpt, pro slovník bylo použito asi tři a půl milionů lístků. Oxfordský slovník (*The Oxford English Dictionary on Historical Principles*, 1933) měl v té době základnu asi pět milionů excerpt.

Na uvedený systém excerptce bylo navázáno i v následujících obdobích, kdy po druhé světové válce proběhla přeměna Kanceláře Slovníku jazyka českého v Ústav pro jazyk český. V 50. letech minulého století se začalo pracovat na koncepci dalšího výkladového slovníku češtiny, *Slovníku spisovného jazyka českého*

(SSJČ, 1960–1971), a s ním souvisel i vznik nové excerptní řady lístkového NLA, tzv. **2. materiálu**. Excerpta pocházejí zejména ze zdrojů vzniklých po r. 1945 až do 2. poloviny 50. let. Objevila se také potřeba vyčlenit slovní zásobu související se změnou politického režimu a s nástupem komunistické éry (*brigádnice, marxismus-leninismus, mládežník* aj.).

3. materiál lístkového NLA (z konce 50. let do 1. poloviny 60. let 20. století) byl oddělen od druhé vrstvy zejména kvůli novému směru v excerptování orientovanému hlavně na publicistický styl. Později byl tento lexikální materiál využit především pro jednosvazkový výkladový slovník spisovného jazyka, *Slovník spisovné češtiny pro školu a veřejnost* (SSČ, 1. vydání 1978).

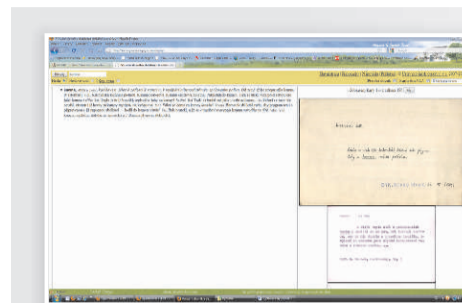


Excerptní listek – 2. materiál lístkového NLA

4. materiál NLA (z konce 60. let do 1. poloviny 70. let 20. století) byl od třetí lexikální vrstvy oddělen z praktických důvodů; na základě 4. materiálu se připravovaly doplňky pro SSJČ. Kromě toho je zde zachycena i excerptce z období uvolňování politického režimu, tj. z 2. poloviny 60. let minulého století. V té době byl excerptován hlavně oficiální proud literatury a probíhala doplňková excerptce starších autorů. Do této vrstvy se však podařilo zahrnout i díla autorů, kteří v pozdějším období nebyli politickým režimem akceptováni – např. od J. Škvoreckého byla vyexcerptována *Legenda Emöke* a *Zbabeleci*, od L. Vaculíka *Sekyra*.

5. materiál NLA (z 1. poloviny 70. let 20. století až do roku 1985) byl také využit při sestavování SSČ a měl sloužit i pro vznik *Slovníku lexikálního standardu*. Tento slovník však zůstal ve fázi přípravných prací a je zachován jen heslář bez uvedení autorství a datace. V druhé polovině 80. let nastala krátká pauza v excerptování, protože se pracovalo na *Akademickém slovníku cizích slov* (1995), k němuž probíhala speciální excerptce, ale ta zůstala nezačleněna do hlavní části lístkového NLA.

6. materiál lístkového NLA zahrnuje období od konce 80. let 20. století do roku 1991 a je poslední vrstvou lístkového lexikálního archivu. V 6. vrstvě se vyskytují kromě jiných záznamů příklady nové lexikálně-sémantické spojitelnosti, nová slova a nová pojmenování (*biopole, brifing*). Tato vrstva NLA sloužila také jako výchozí zdroj pro publikaci *Co v slovnících nenajdete. Novinky v současné slovní zásobě* (1994).



Příruční slovník jazyka českého a databáze lístkového NLA

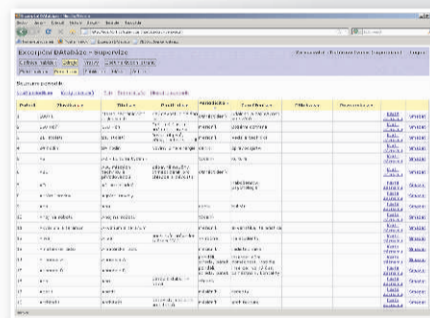
V roce 2007 bylo v Ústavu pro jazyk český AV ČR, v. v. i., dokončeno **skenování základní části lístkového lexikálního archivu** (materiálů 1.–6.), obsahující **8 696 850 excerpt**. Snímky excerptních záznamů v elektronické podobě jsou zařazeny do databáze. Vyhledávací programy jsou plně funkční pro různé internetové prohlížeče a aplikace je přístupná na webových stránkách ÚJČ AV ČR, v. v. i., z adres <http://bara.ujc.cas.cz/psjc> a <http://lexiko.ujc.cas.cz> (v rubrikách *Databáze heslářů* a *Odkazy*). **Unikátní je zejména propojení této databáze s elektronickou podobou Příručního slovníku jazyka českého**, umožňující práci s oběma zdroji zároveň.

2. Nový elektronický lexikální archiv

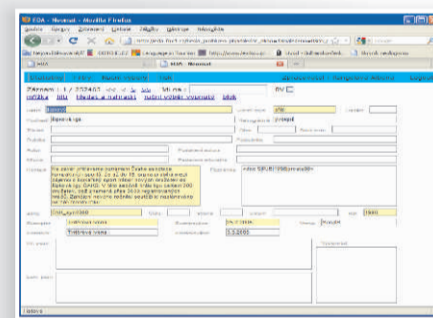
Na počátku 90. let 20. století přispěly výrazné změny v oblasti techniky a komunikace i ke změnám v postupech lingvistické práce a došlo rovněž k přechodu na efektivnější elektronickou excerptci. **Od listopadu 1991 je lexikální materiál ukládán elektronicky s využitím speciálních excerptních programů**. Vzhledem k rostoucí technické náročnosti lexikografické práce provedl nedávno excerptní úsek převod programu *WinHesla2* na databázový systém, využívající internetové technologie, s názvem *EDA*. Elektronický lexikální archiv navazuje na lístkový systém excerptce, ale díky počítačovému zpracování se otevřely nové možnosti pro sběr jazykového materiálu.

Nový elektronický lexikální archiv zahrnuje v současné době dvě databáze: **excerptní databázi**, obsahující neologický lexikální archiv (*Archiv 1*) a archiv lexikální dynamiky (*Archiv 2*), a **excerptní oborovou databázi (ODE)**.

Vznik **neologického archivu** úzce souvisel s lexikografickým popisem nových slov, tzv. **neologismů** (publikace: *Nová slova v češtině. Slovník neologismů 1 a 2*). **Archiv lexikální dynamiky** navazuje na neologickou excerptci, ale byl oddělen od předchozího materiálu, neboť je spojen s realizací výzkumného záměru (VZ) *Vytvoření databáze lexikální zásoby českého jazyka počátku 21. století* (2005–2011), a zachycuje jevy lexikální dynamiky v širším smyslu (např. nové významy slov, nové víceslovné lexikální jednotky, změněnou spojitelnost slov, jiné stylistické hodnocení ad.). K 31. 12. 2010 bylo v elektronické excerptní databázi uloženo 252 539 excerptních záznamů.



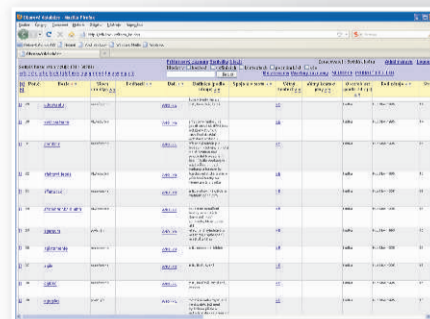
Excerptní databáze EDA



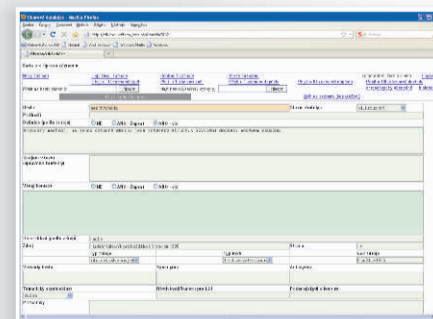
Formulář pro zápis do excerptní databáze EDA

První část neologického archivu – **databáze Neomat** – je zpřístupněna nejširší veřejnosti na adrese: www.neologismy.cz.

V rámci výše uvedeného VZ byla v roce 2009 navržena a zprovozněna samostatná **excerptní oborová databáze (ODE)**. Zatím je v ní vloženo přes 60. tis. záznamů, zachycujících výrazivo z různých oborů a oblastí lidské činnosti. Shromážděný jazykový materiál se tak postupně rozšiřuje o slovní zásobu z učebnicových, odborných a popularizačních textů.



Excerptní oborová databáze ODE – souhrnná tabulka



Formulář pro zápis do excerptní oborové databáze ODE

Lístková část NLA byla postupně využita při tvorbě všech akademických výkladových slovníků češtiny – *Příručního slovníku jazyka českého*, *Slovníku spisovného jazyka českého* a *Slovníku spisovné češtiny pro školu a veřejnost*, lingvisté z ní čerпали i při tvorbě slovníků překladových a speciálních. Ovšem nesloužila a neslouží jen pro lexikografickou praxi, na jejím základě vznikla také řada dalších domácích i zahraničních vědeckých prací. Může být pramenem pro poznání vývoje pravopisu, gramatiky, stylistiky a pro další lingvistická badání. Do budoucna je plánován přesun tohoto velmi cenného jazykového materiálu, který má dnes mimo lingvistické také kulturně-historickou hodnotu, do Archivu AV ČR. Nový elektronický lexikální archiv NLA by měl být nadále promyšleně budován pro stávající i budoucí lexikografické výstupy.