

## QUELQUES MOTS SUR LA LEMMATISATION

### Points principaux de l'exposé :

- 1 Objectifs de la lemmatisation
- 2 Méthodes de la lemmatisation
  - 2.1 Listes des lemmes et de leurs formes / R. Busa, Gallarate, Pisa
  - 2.2 "Radicaux" et morphèmes grammaticaux "désinences" / L. Delatte, Liège
  - 2.3 Caractérisation par groupes, caractérisation en étapes  
Exploitation des index alphabétiques, y compris index inverses,  
et du rangement dans l'ordre du texte / Institut pour la Langue  
Tchèque, Prague
- 3 Problèmes de l'homographie
- 4 Questions de l'individualisation des lemmes, variantes et doublets
  - 4.1 Principes de l'individualisation des lemmes et détermination de  
leurs formes premières
  - 4.2 Doublets de dérivation
  - 4.3 Doublets lexicaux
  - 4.4 Variantes et divergences orthographiques
  - 4.5 Doublets et variantes vocaliques
  - 4.6 Doublets et variantes stylistiques
  - 4.7 Noms propres
  - 4.8 Mots étrangers en citations

Mesdames, Messieurs,

Permettez-moi de vous présenter quelques observations relatives à la lemmatisation. Je me propose de les résumer rapidement sur la base de notre modeste expérience, recueillie à l'étude d'un texte tchèque moderne.

Je dois dire, pour commencer, qu'à la rédaction de mon rapport je n'ai pas encore disposé du très utile recueil "Actes du Séminaire international sur le Dictionnaire latin de la machine" /1/, sur les objectifs de la lemmatisation, s'occupant des procédés appliqués à l'analyse du texte et spécifiant deux méthodes de lemmatisation. J'avais en effet l'intention d'attirer votre attention en premier lieu sur la nécessité de porter à la connaissance de tous les intéressés les procédés de lemmatisation, les problèmes concrets se rattachant à l'étude des textes et l'expérience qui en résulte.

#### OBJECTIFS DE LA LEMMATISATION

Dans le traitement mécanographique, la lemmatisation a pour but d'associer chaque unité du texte au lemme respectif, c'est-à-dire de la caractériser au point de vue morphologique, et d'en établir la forme première, représentative, le lemme au sens restreint du terme.

Pris dans sa conception générale, le lemme représente non seulement cette forme première, mais potentiellement aussi le paradigme complet du mot-type.

Les caractéristiques morphologiques permettent de procéder, sur leur base, à l'analyse du texte.

On entend par lemmatisation le procédé technologique d'une certaine description lexicographique et en même temps le résultat obtenu de cette façon. Considérée comme procédé technologique, la lemmatisation est d'un contenu variable, comme la technologie même peut varier et varie en fait. Le procédé retenu est motivé par les moyens disponibles, d'une part, et surtout par les objectifs à réaliser, de l'autre; s'il s'agit d'une documentation prévue pour l'analyse générale du texte ou seulement d'un travail lexicographique de caractère descriptif. Au sens le plus restreint du terme, la lemmatisation se réduit à l'établissement des formes premières, satisfaisant les buts essentiels de la lexicographie, la rédaction des index et des concordances.

## 2 METHODES DE LA LEMMATISATION

Dans la pratique actuelle du traitement mécanographique, la définition morphologique des formes et leur association aux lemmes respectifs se font en principe de deux façons : on dresse les listes des lemmes et de leurs formes /2.1/ ou bien les listes des radicaux ainsi que celle des morphèmes, des désinences /2.2/. Nous avons en partie combiné ces deux méthodes et réalisé la lemmatisation en étapes, à l'emploi simultané du rangement inverse /2.3/.

- 2.1 La première méthode, première en fait du point de vue historique c'est-à-dire nettement antérieure à toutes les autres, est basée sur les listes de lemmes et de leurs formes. A la rédaction de ces listes, on prend pour point de départ l'inventaire lexicographique, énumérant ces formes nominatif. La méthodologie respective est due au Professeur P. R. Busa et a

été décrite dans les études qu'il a publiées /2/. La rédaction de ces "index documentaires de confrontation" des lemmes dont chacun porte son numéro d'identification, donc des listes des formes premières et formes de flexion, est basée sur le Dictionnaire latin de Forcellini /3/.

2.2 La seconde méthode de lemmatisation n'a pas recours à des listes de lemmes, spécifiant systématiquement et nominativement toutes les formes selon le dictionnaire. La sélection des radicaux, parties invariables des mots pouvant recevoir un certain répertoire de morphèmes grammaticaux, permet d'établir, au fur et à mesure, le plan de confrontation dicté par l'étude du texte. Cette méthode est réalisée dans le Laboratoire de Liège, sous la direction du Professeur Louis Delatte. Elle s'appuie non seulement sur un dictionnaire, en l'occurrence également sur le Dictionnaire de Forcellini, mais aussi, et surtout, sur les descriptions grammaticales du système morphologique. A l'encontre de la première méthode, la seconde permet de réduire la capacité de la mémoire externe, mais augmente le volume d'opérations à exécuter par la machine et celui des préparatifs /4/.

De plus, elle donne naissance à certains problèmes nouveaux, tels que homographie des radicaux et morphèmes grammaticaux, nécessité d'enregistrer et d'agrèger les éléments allomorphes, etc. Je rappelle que je pars de l'idée d'application de cette méthode à la spécification des unités constituant le système morphologique tchèque où les listes des cas irréguliers ne pourraient pas couvrir toutes les alternances de lettres dans les radicaux des mots, à la flexion.

2.3 Nous avons été obligés de tenir en ligne de compte surtout les raisons

d'économie, soit spécialement le fait que nous ne disposions pas de machines appropriées. Cet état de choses nous a conduit à mettre au point un procédé particulier, combinant les deux méthodes précédentes. Le travail de la machine est en l'occurrence assez sensiblement complété d'interventions humaines, surtout dans les phases intermédiaires. Notre tâche consistait à établir, à partir du texte examiné, un index de concordances, répertoriant les diverses formes des mêmes mots. Nous nous sommes donc décidés à nous procurer rapidement les formes premières, en laissant pour plus tard le traitement et l'enregistrement de l'inventaire morphologique pour la majorité des unités. Nous avons dû chercher un régime d'économie.

Après l'enregistrement mécanographique du texte, nous avons rangé par ordre alphabétique le fichier lexical et rédigé un index de fréquences. Partant de celui-ci et des mots les plus fréquents, nous avons ensuite séparé les mots synsémantiques, inflexibles.

Ces mots synsémantiques et inflexibles reçoivent leur caractérisation immédiatement et on complète, par reproduction, le lemme identique à la forme. De même, les mots très fréquents, sans cas homographiques, sont caractérisés sans délai et reçoivent leur lemme.

Une fois terminée la caractérisation des mots très fréquents, l'inventaire est rangé alphabétiquement, par ordre inverse, et on établit en même temps l'index inverse de toutes les différentes formes /selon l'index de fréquences/.

D'après l'index inverse des formes, comprenant aussi les données de fréquence, on sépare, dans l'index inverse "complet", les groupes de mots, dont les désinences permettent de déterminer sans équivoque une catégorie grammaticale et de fournir le lemme respectif, soit par reproduction, soit à l'aide de programmes établis, réalisant les alternances des lettres de radical et les changements de désinences.

Les unités qui sont restées sans lemme sont rangées selon le texte, possèdent-elles ou non une caractérisation grammaticale, complète ou partielle. L'ordre du texte fournit les formes premières d'une manière rapide, sans caractérisation morphologique.

En fin de cette étape, il est possible d'établir un index "lemmatisé" à formes concentrées en lemmes, mais rangées par ordre alphabétique. Pour obtenir l'ordre systématique, on ne dispose pas en attendant d'appareil nécessaire, soit caractéristiques en code, soit ordre systématique de confrontation sous forme d'index de lemmes et de leurs formes. Pour plus de détails, se référer au spécimen d'index à l'annexe, présentant les formes concentrées en lemmes et confrontant deux éditions du texte : Petr Bezruč : Slezské písně, éditions de 1957 et de 1928 /4/.

Le contrôle et la mise au point des caractéristiques morphologiques ne sont réalisés qu'à l'étape suivante, constituant une phase indépendante.

Avec l'index inverse de tous les mots du texte et celui de toutes les différentes formes, on obtient aussi l'index inverse de lemmes. Les index inverses permettent d'adjoindre les caractéristiques aux groupes complets

de mots, selon les éléments de dérivation, tout en présentant l'inventaire pour la description et la séparation de l'homographie grammaticale, selon les morphèmes-désinences de l'index de formes /5/.

### 3 PROBLEMES DE L'HOMOGRAPHIE

Une des entraves les plus graves à la rapidité et à la sûreté du traitement automatique des textes est l'existence d'homographes. En effet, le fait qu'une seule forme est commune à deux lemmes est assez fréquent. Il ne s'agit pas de distinguer seulement les cas d'homonymie lexicale; l'essence du problème est l'homonymie "grammaticale", les morphèmes homographiques de caractère grammatical. L'homographie existe déjà dans le paradigme d'unités lexicales, répondant à l'économie de la langue et de son système. L'homographie apparaît aussi dans les radicaux, dans le morphème de base; par exemple, le mot "náš" est : 1) le pronom possessif, 2) le radical des mots composés : na-náš-eti. Pour l'inventaire des radicaux, il faut au contraire assembler les radicaux allomorphes. Donnons, à titre d'exemple, les morphèmes verbaux de base : nos, náš, nés dans les verbes : nos-iti, s-náš-eti, nés-ti.

Il est vrai qu'on n'ignore pas les problèmes se rattachant aux homographes le Professeur P. R. Busa /voir la note 2/ a attiré l'attention sur les difficultés qui leur sont propres déjà en 1961; mais il n'en est pas moins vrai que leur analyse ne reçoit que des soins théoriques relativement réduits. Le plus souvent, les caractéristiques distributives ne portent que sur des spécimens bien sélectionnés et non sur un texte complet.

#### 4 QUESTIONS DE L'INDIVIDUALISATION DES LEMMES, VARIANTES ET DOUBLETS

Un problème non moins délicat, ou au moins encore ouvert, est la détermination ou le choix de la forme première, du lemme au sens restreint du terme. L'étude de notre premier texte a pour objet un livre de poèmes dont les différentes publications, couvrant une cinquantaine d'années, nous nous sommes proposé de confronter lexicalement. Ce fait nous a conduit à la nécessité de préciser ce que l'on entend par les termes : variante lexicale et doublet lexical, et d'établir des règles pour leur traitement. Il serait en effet inutile, sinon gênant, de séparer en lemmes les différentes formes orthographiques des éditions consécutives, ne traduisant souvent que la pratique de l'époque.

Il faut toutefois fixer, comment traiter les doublets orthographiques, stylistiques, de dérivation et lexicaux, pour distinguer ceux qui ne sont que de simples variantes et ceux qui sont des doublets effectifs, en vue d'établir ces derniers comme lemmes autonomes.

- 4.1 L'individualisation d'un lemme part du principe morphologique. Ce postulat a décidé sans équivoque du traitement des noms et des adjectifs verbaux, des comparatifs et superlatifs des adjectifs et adverbess, des formes verbales négatives, des mots étrangers.

Nous tenons chaque nom verbal pour nom et non pour une forme du verbe respectif, bien que le niveau de lexicalisation soit en l'occurrence variable.

De même les adjectifs verbaux, issus du gérondif présent, du participe passif ou du participe passé, forment tous des lemmes autonomes, même s'ils apparaissent en forme nominale. On ne prend pour partie ni des formes verbales composées ni le participe passif, constituant avec la copule l'attribut verbal dénominatif.

Les comparatifs et les superlatifs des adverbes sont rangés sous les positifs respectifs, si leur formation est régulière. Si elle est irrégulière, on fait appel au Dictionnaire de la langue tchèque littéraire /6/ : s'ils figurent dans le Dictionnaire comme mots autonomes, le lemme respectif reçoit également une place indépendante; si ce n'est pas le cas, ils sont rangés sous le positif, avec référence éventuelle.

Les formes négatives du verbe figurent sous les formes positives; la même règle est appliquée à l'adverbe *nelze - lze*. Les formes négatives des autres mots constituent des lemmes autonomes /adjectif : *nedočknutelný*; nom : *nepesina*, etc./.

- 4.2 Sont aussi lemmes indépendants tous les mots dérivés. On considère donc comme doublets lexicaux constituant des lemmes parallèles aussi les doublets de dérivation. Par exemple, les parallèles des diminutifs dérivés, les adjectifs parallèles issus des noms propres, surtout de caractère local, les adjectifs possessifs, etc.
- 4.3 Il n'y a pas de doute que les doublets lexicaux ont leur propre existence. En cas de doublets morphologiquement difficiles à distinguer, lorsque par exemple la désinence de cas n'indique pas suffisamment le lemme /ainsi, le

génitif “země” peut s’associer aussi bien au nominatif “zem” qu’au nominatif “země”/, on constitue des lemmes pour ainsi dire “hybrides”; par exemple zem/země, rámě/rameno, cos/cosi, víc/více.

4.4 Indiscutable est au contraire l’association des variantes et divergences orthographiques, et ce sous la forme codée, pour autant qu’elle existe. Si l’usage actuel est inconnu ou s’il n’est pas régularisé, la première rencontre du mot domine la forme du lemme. Cela est vrai surtout pour les groupes initiaux vz/z, sz/s : vzpurný spurný; zsinalý sinalý, aussi bien que pour la régularisation des voyelles i et y /cymbál-cimbál/, pour la simplification des lettres doublées, etc.

En ce qui concerne la confrontation des éditions consécutives, il est peut-être indiqué de prendre ces faits en considération même si les éditions antérieures n’ont pas de doublets, mais seulement une orthographe conforme à l’usage ancien. Cela revient à dire qu’il convient de prévoir des lemmes pour les formes orthographiques courantes à l’époque et d’ajouter une référence, s’il y a lieu.

4.5 Dans le domaine des variantes vocaliques, la situation est fort compliquée. Les alternances vocaliques sont le thème des doublets de dérivation, orthographiques et stylistiques. A ces cas, on applique principalement le critère individuel de divers types de parallèles. Pour les parallèles simples à alternances vocaliques, nous avons souvent adopté les lemmes “hybrides”, associant les variantes. Cela est vrai en premier lieu pour les différences en quantité des voyelles radicales. Si la forme à quantité graphiquement marquée est littéraire, tandis qu’elle est du patois en forme sans quantité,

le mot est rangé sous le lemme littéraire à quantité marquée : *dymka* et *dýmka* sous le lemme *dýmka*; *steblo* et *stéblo* sous le lemme *stéblo*. Si les deux formes sont littéraires, on établit un lemme "hybride" : *jizlivý/jžlivý*; *velký/veliký*, etc.

- 4.6 Une attention particulière et une étude poussée sont à consacrer aux variantes et doublets à thème stylistique, qu'il s'agisse d'une parallèle à deux formes faisant thème stylistique ou d'une parallèle des formes thématique et neutre.

A l'étude de notre texte, on avait surtout affaire aux mots et formes de mots pris du patois de la langue littéraire et d'autre part aux mots d'une fréquence restreinte, mots rares, désuets et d'un usage limité /littérature, poésie/.

Le plus souvent, l'emploi de ces mots non neutres a une certaine fonction. Dans les éditions antérieures, il faut distinguer les mots désuets, correspondant à la norme lexicale de l'époque, et les mots qui ont une certaine fonction, étant assez peu fréquents. Aussi, a-t-on laissé de côté les mots désuets, par exemple : *hetman*/de nos jours : *hejtman*/, *dragon*/ en langue littéraire : *dragoun*/, *vědma*/ en langue littéraire : *vědma*/, *knut*/ en langue littéraire : *knuta*/, etc. Les mots du patois, sans pendant littéraire, constituent des lemmes indépendants /*četyna*, *potka*/ aussi bien que les mots qui ont un pendant littéraire, ou pourraient l'avoir, mais sont employés dans une acception tout à fait différente /*iskerka* = *sedmikráska* tandis que *jiskerka* est le diminutif littéraire de *jiskra*/. On voit que dans ces cas la lemmatisation dépend du sens. Mais on ne respecte pas la polysémie.

Les unités polysémiques ne sont pas distinguées. Ce n'est que pour les noms propres, de caractère local ou personnel, que le sens décide encore de l'homonymie.

4.7 La lemmatisation des noms propres et de leurs parties constitue un domaine tout spécial. Pour les noms de localités, ne différant de la forme littéraire actuelle que par la quantité des voyelles, on a adopté comme lemme la forme littéraire : donc lemme *Frýdek*, forme *Frydek*; lemme *Micháلكovice*, forme *Michalkovice*. S'il y a dans le texte une forme qui n'existe plus de nos jours, elle reçoit son propre lemme. Dans un cas pareil, on indique cependant la forme actuelle comme référence : lemme *Střebovice*, forme *Střebovice*; référence *Třebovice*, voir *Střebovice*.

Les noms communs, faisant en même temps office de noms propres ou de leurs parties, n'ont pas de lemmes. Un seul lemme associe ainsi les formes aux lettres initiales alternatives. Mais en revanche, on range comme lemme le féminin d'adjectifs */Suchá, Krásná/* qui n'est pas rangé sous le masculin */suchý, krásný/*.

4.8 Les noms étrangers utilisés en citations, donc non les noms empruntés à la langue étrangère, ne sont pas en général lemmatisés; on prend comme lemme la forme du texte original. Par exemple, dans les vers allemands : "Lege dar, wo mährisch Blut hier - vielleicht geh ich", les mots : lege et geh sont répertoriés comme lemmes tels quels, et non comme infinitifs allemands : legen, gehen. De la locution latine : Praga caput regni, le mot : regni est pris comme lemme, et non le nominatif : regnum. Font exception les textes à plusieurs formes qu'il faut ranger sous un seul lemme;

dans ce cas, on prend la forme première de la langue étrangère.

Les machines exécutent le traitement du texte rapidement, avec précision et à l'aide d'une mémoire infallible; elles n'en sont pas cependant moins des instruments. Sans recours à l'intelligence humaine, nul travail n'est réalisable, ni les opérations auxiliaires, telles que enregistrement et documentation lexicographique du texte.

J'ai tâché de vous expliquer rapidement quelques solutions que nous avons appliquées.

Institut de la Langue tchèque  
Académie des Sciences tchécoslovaques

Jitka ŠTINDLOVÁ

## NOTES

- /1/ Revue, 1969, n<sup>os</sup> 2 et 3, Liège; Calcolo, Supplément n<sup>o</sup> 2, vol. V, 1969, Roma, Pisa.
- /2/ Rappelons seulement les études les plus importantes :  
*Zusammengefasste Darstellung der Erfahrungen des Centro per l'Automazione dell'Analisi Letteraria des Aloisianum, Gallarate (Varese) Italien*, Kolloquium Maschinelle Methoden der literarischen Analyse und der Lexikographie, Tübingen, 24-26 Nov. 1960, pp. 6-26.  
*Les travaux du Centro per l'Automazione dell'Analisi Letteraria de Gallarate* in Cahiers de Lexicologie, vol. 3, 1961, Paris, Didier 1962, pp. 64-68.  
*L'analyse des homographes et des formes flexionnelles* in Cahiers de Lexicologie, vol. 3, 1961, Paris, Didier 1962, pp. 134-139.  
*Un lexique latin électronique. Les Machines dans la Linguistique.* Colloque Intern. sur la mécanisation et l'automation des recherches linguistiques. Academia, éd. de l'Acad. Tchecoslovaque des Sciences, Prague, 1968, pp. 251-269.
- /3/ Forcellini : Lexicon Totius Latinitatis.
- /4/ Rappelons les Actes du séminaire international sur le dictionnaire latin de machine, voir note /1/ et quelques études fondamentales :  
L. Delatte, Programme d'analyse morphologique et syntaxique du latin, Revue, 1965, n<sup>o</sup> 1, pp. 57-62.  
A. Bodson, S. Govaerts, Codification d'un texte latin sur cartes mécanographiques IBM-80 colonnes, Revue, 1966, n<sup>o</sup> 1, pp. 1-46.  
A. Bodson, E. Evrard, Le programme d'analyse automatique du latin Revue, 1966, n<sup>o</sup> 2, pp. 17-46.  
A. Bodson, Le programme d'analyse automatique du latin, Les machines dans la linguistique, Academia, Prague 1968, pp. 117-130.

/5/ Voir l'annexe :

Le recueil "Slezské písně" de Petr Bezruč paraît être un ouvrage très convenable pour le traitement mécanographique, sur la base duquel on pourra élaborer la méthode d'établissement des index et vocabulaires et celle de comparaison des textes.

Petr Bezruč /1867-1958/ est entré dans la littérature au commencement du siècle par des vers très impressionnants sur l'oppression nationale et sociale du peuple tchèque vivant dans la Moravie du Nord. Il a publié un recueil de poèmes /1909/, pleins d'esprit de révolte, qui se sont répandus dans un grand nombre d'éditions, souvent modifiées par l'auteur, et qui ont été traduits en beaucoup de langues étrangères. La poésie de Bezruč est toujours actuelle non seulement par ses sujets, mais particulièrement par sa valeur artistique.

/6/ Je me propose de consacrer plus tard une étude spéciale aux possibilités d'emploi des index inverses /mots, formes, lemmes/ dans la caractérisation grammaticale des unités de textes. Certains renseignements sur ce sujet sont à trouver dans les publications suivantes :

*Sur le classement inverse des mots et sur ce qu'on appelle "dictionnaire inverse"*, Cahiers de lexicologie (Besançon) 2, 1960, pp. 79-86.

*Retrográdní slovníky. Slovo a slovesnost* (Praha) 21, 1960, pp. 270-272. *Po druhé o retrográdních slovnících. Slovo a slovesnost* 27, 1966, pp. 370-374.

/7/ La lemmatisation doit toujours chercher un appui dans la lexicographie. En l'occurrence, nous faisons appel au Dictionnaire mentionné rédigé, sous la direction du Professeur B. Havránek, par un groupe d'auteurs de l'Institut pour la langue tchèque près de l'Académie tchécoslovaque des Sciences, et publié par les soins de l'Academia 1960-1970.

A n n e x e s :

1. Spécimen de l'index de fréquence relatif à l'ouvrage de Petr Bezruč :  
Slezské písně.
2. Spécimen de l'index avec confrontation, sous les lemmes respectifs, de  
deux éditions du texte : Slezské Písně de Petr Bezruč / 1957 et 1928/.

a	459	2se	28	píseň	17
v	417	lby	27	ruce	17
l se	309	můj	27	či	16
na	257	si	27	děvucha	16
z	245	poli	26	není	16
jak	180	.bůh	25	noc	16
do	146	hledí	25	pán	16
jsem	145	kdy	25	pro	16
lco	120	2má	25	rád	16
tak	106	tě	25	těch	16
l je	101	zem	25	.beskyd	15
za	101	mnou	24	časem	15
mi	93	žití	24	.dombrová	15
s	93	cizí	23	lhoře	15
pod	90	ale	22	i	15
to	90	2až	22	jí	15
po	89	jsi	22	markýz	15
když	88	.těšina	22	národ	15
k	79	než	21	té	15
já	64	tož	21	toho	15
tam	64	.gero	20	2tu	15
kdo	58	krev	20	ženy	15
mne	57	mé	20	bych	14
jen	52	mu	20	bylo	14
nad	51	oči	20	bys	14
l nech	49	.ostravice	20	děvuchy	14
že	49	srdce	20	dva	14
ve	48	sto	20	hlavu	14
před	43	jim	19	své	14
kde	42	slunce	19	ta	14
od	42	šel	19	tvé	14
l ty	42	už	19	vlasy	14
ten	41	černé	18	však	14
jsou	37	jeden	18	děvče	13
při	37	jej	18	dnes	13
nás	36	mně	18	ještě	13
ze	36	nocí	18	krve	13
nám	35	řečí	18	lidu	13
byl	33	život	18	.magdonova	13
2ty	32	den	17	měl	13
země	32	duši	17	muž	13
jde	31	dvě	17	naše	13
l jednou	31	.frýdku	17	pole	13
l ti	31	ji	17	sám	13
jako	30	jich	17	školy	13
bez	29	kopu	17	tomu	13
jest	29	l má	17	vás	13
přes	29	my	17	žije	13
u	29	neq	17	beze	12
bude	28	o	17	černý	12

děti	12	lid	10	jdou	8
kam	12	lidí	10	2je	8
květ	12	.maryčko	10	lkolem	8
moje	12	národa	10	láska	8
ne	12	nikdy	10	leží	8
nebe	12	očí	10	.lysou	8
on	12	ohně	10	.maryčka	8
rež	12	polské	10	mluví	8
rukou	12	roků	10	moravský	8
řeč	12	sedí	10	neb	8
školu	12	snad	10	neseš	8
táhne	12	stál	10	nevím	8
.těšínem	12	tisíců	10	oko	8
vy	12	víc	10	osud	8
žena	12	všecko	10	.polské	8
.bernard	11	zemi	10	práci	8
budeš	11	zní	10	proč	8
daleko	11	zvoní	10	prvý	8
ho	11	žal	10	půjdu	8
2jeho	11	.beskydem	9	rty	8
jenž	11	běží	9	sedm	8
koně	11	byla	9	smích	8
moravská	11	dá	9	sobě	8
ní	11	juž	9	starý	8
pluh	11	každý	9	svět	8
ráz	11	kdož	9	svou	8
svých	11	kmen	9	syn	8
synek	11	let	9	tich	8
.šla	11	mojí	9	tobě	8
2ti	11	ním	9	.tošonovic	8
ticho	11	.ondro	9	zemí	8
tvář	11	přijde	9	žid	8
ví	11	rudý	9	.žor	8
viděl	11	sbohem	9	aby	7
vždy	11	tisíc	9	bo	7
žil	11	tři	9	břehu	7
.bože	10	ven	9	čelo	7
bratře	10	vidím	9	čí	7
buď	10	voda	9	či	7
dávno	10	zde	9	dobrý	7
dědiny	10	1až	8	dobře	7
dej	10	ber	8	doly	7
děvucho	10	bílé	8	domky	7
dým	10	dal	8	duše	7
hory	10	ldolů	8	grof	7
chce	10	2dolů	8	hej	7
chloup	10	goral	8	hlavy	7
jejich	10	hlava	8	hled	7
jsme	10	hor	8	horách	7
kalina 3	10	chodil	8	horemi	7

	1957	1928	217
MELECKÁ	1 80 022 01		
MELODIE	2*	2*	
MELODIE	09 000 02	09 000 02	
MELODIE	2 09 012 04	2 09 012 04	
V MENITI	2*	2*	
V MENI	1 68 012 04	1 68 012 04	
V MENIM	1 08 013 06	1 08 013 06	
VV MERIC	1*	1*	
VV MERICI	1 54 006 01	1 54 006 01	
VV MERITI	1*	1*	
VV MERIL	1 38 010 07	1 38 010 07	
V MESIC	5*	5*	
V MESIC	30 010 05	30 010 05	
V MESIC	2 31 013 02	2 31 013 02	
V MESICE	65 041 02	65 041 02	
V MESICE	2 77 030 04	2 77 030 04	
V MESICI	1 52 024 02	1 52 024 02	
V MESTECKO	1*	1*	
V MESTECKA	1 07 009 01	1 07 009 01	
V MESTO	17*	17*	
V MEST	1 63 003 04	1 63 003 04	
V MESTA	08 021 03	08 021 03	
V MESTA	19 004 05	19 004 05	
V MESTA	41 012 07	41 012 07	
V MESTA	48 027 05		
V MESTA	52 006 04	52 006 04	
V MESTA	60 002 05	60 002 05	
V MESTA	7 70 002 02	6 70 002 02	
V MESTE	40 013 02	40 013 02	
V MESTE	2 52 098 06	2 52 098 06	
V MESTECH	1 02 007 05	1 02 007 05	
V MESTEM	48 030 01	48 030 01	
V MESTEM	2 70 013 03	2 70 013 03	
V MESTO	29 021 04	29 021 03	
V MESTO		48 027 04	
V MESTO	2 63 011 06	3 63 011 06	
V MESTU	08 029 07	08 029 07	

Spécimen 2

	1957	1928	218
V MĚSTU	2 65 031 06	2 65 031 06	
V MĚSTSKÝ	1*	1*	
V MĚSTSKÝM	1 02 016 04	1 02 016 04	
METELICE	1*	1*	
METELICE	1 77 011 01	1 77 011 01	
METLA	1*	1*	
METLOU	1 07 019 05	1 07 019 05	
METLICE	1*	1*	
METLICE	1 33 014 03	1 33 014 03	
METROPOL	1*	1*	
METROPOLI	1 63 003 05	1 63 003 05	
MEZ	3*	3*	
MEZ	1 66 015 07	1 66 015 07	
MEZICH	14 021 03	14 021 03	
MEZICH	2 73 002 02	2 73 002 02	
MEZI	6*	6*	
MEZI	04 005 02	04 005 02	
MEZI	15 008 02	15 008 02	
MEZI	27 047 04	27 047 04	
MEZI	28 002 02	28 002 02	
MEZI	56 030 02	56 030 02	
MEZI	6 65 005 03	6 65 005 03	
V MHOURITI	1*	1*	
V MHOURI	1 67 006 03	1 67 006 03	
MIDAS	1*	1*	
MIDAS	1 18 021 05	1 18 021 05	
MIH	1*	1*	
MIHU	1 08 023 07	1 08 023 07	
MIHNOUI	3*	3*	
MIHL	1 11 007 05	1 11 007 05	
MIHLY	1 83 004 02	1 83 004 02	
MIHNOU	1 39 003 04	1 39 003 04	
MICHÁLKOVICE	2*	2*	
MICHÁLKOVICE	1 36 000 01	1 36 000 01	

	1957	1928	219
·MICHALKOVICICH	1 59 031 07	1 59 031 07	
MILE	4*	4*	
MIL	43 035 05	43 035 05	
MIL	2 43 062 05	2 43 062 05	
MILI	50 068 06	50 068 06	
MILI	2 66 015 03	2 66 015 03	
MILENKA	1*	1*	
MILENKA	1 20 016 01	1 20 016 01	
MILNIK	1*	1*	
MILNIKU	1 08 030 06	1 08 030 06	
MILOSRDNÝ	1*	1*	
MILOSRDNÁ	1 43 077 05	1 43 077 06	
MILOVANI	1*	1*	
MILOVANI	1 08 061 06	1 08 061 06	
MILOVANÝ	1*	1*	
MILOVANÉ	1 67 003 02	1 67 003 02	
MILOVATI	10*	9*	
MILOVAL	01 007 01		
MILOVAL	08 085 01	08 085 01	
MILOVAL	3 10 065 04	2 10 065 04	
MILOVATI	1 69 004 06	1 69 004 06	
MILUJE	07 003 05	07 003 05	
MILUJE	07 015 05	07 015 05	
MILUJE	3 64 011 02	3 64 011 02	
MILUJI	1 72 009 01	1 72 009 01	
MILUJI	1 52 100 02	1 52 100 02	
NEMILUJI	1 06 013 03	1 06 013 03	
MILUJÍCÍ	1*	1*	
MILUJÍCÍ	1 08 082 04	1 08 082 04	
MILÝ	2*	2*	
MILÝ	13 037 01	13 037 01	
MILÝ	2 72 013 03	2 72 013 03	
MINISTR	2*	2*	
MINISTRŮ	1 77 100 02	1 77 100 02	
MINISTRŮM	1 69 005 03	1 69 005 03	
MINOUTI	2*	2*	