

## Tokens

### Interface de acesso ao banco de dados dbtokens

Tokens é uma interface destinada aos administradores do banco de dados dbtokens, criado em MySQL, com as seguintes funcionalidades:

1. Criar, modificar e deletar tabelas;
2. Popular as tabelas manualmente, através do teclado, ou automaticamente, através de scripts em PHP/SQL;

Atualmente, dbtokens contém duas tabelas: nested\_classes e tokens (Fig. 1).

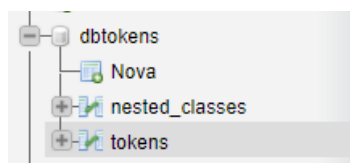


Fig. 1: Tabelas da base de dados dbtokens

As estruturas das tabelas são mostradas nas Figs. 2 e 3:

#	Nome	Tipo	Agrupamento (Collation)	Atributos	Nulo	Predefinido	Comentários	Extra
<input type="checkbox"/>	1 class_id	int(11)			Não	Nenhum		AUTO_INCREMENT
<input type="checkbox"/>	2 name	varchar(200)	latin1_swedish_ci		Não	Nenhum		
<input type="checkbox"/>	3 lft	int(11)			Não	Nenhum		
<input type="checkbox"/>	4 rgt	int(11)			Não	Nenhum		

Fig. 2: Estrutura da tabela nested\_classes

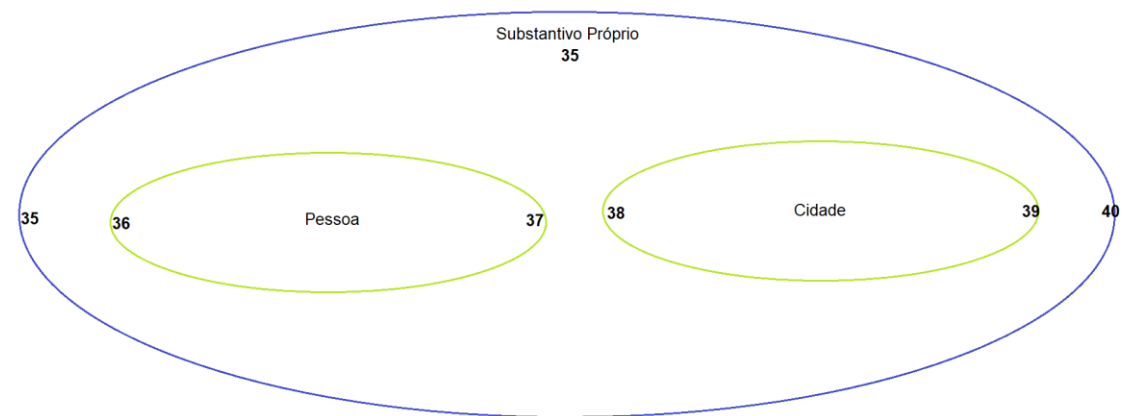
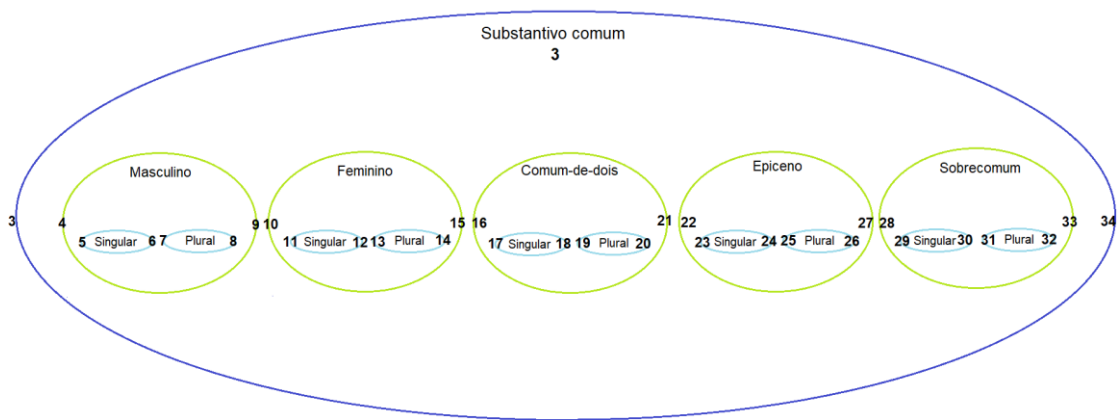
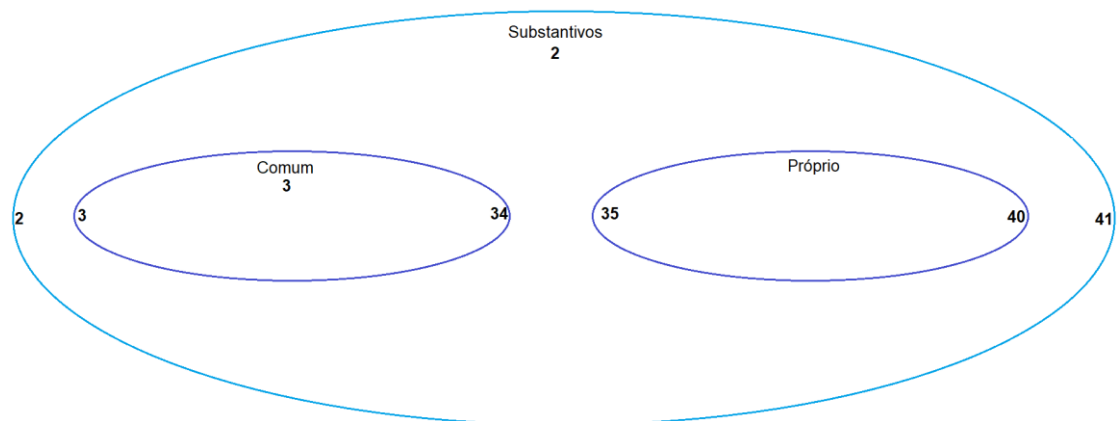
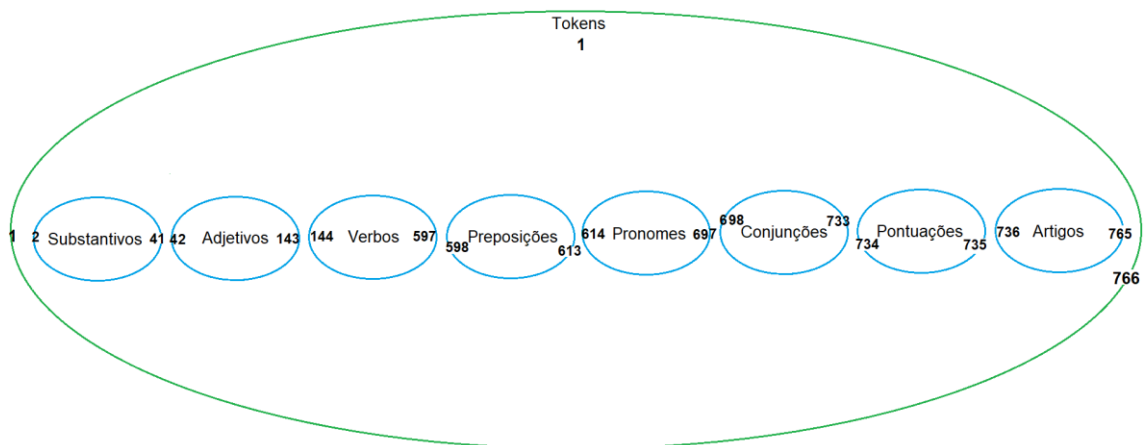
#	Nome	Tipo	Agrupamento (Collation)	Atributos	Nulo	Predefinido	Comentários	Extra
<input type="checkbox"/>	1 token_id	int(11)			Não	Nenhum		AUTO_INCREMENT
<input type="checkbox"/>	2 name	varchar(200)	latin1_swedish_ci		Sim	NULL		
<input type="checkbox"/>	3 class_id	int(11)			Não	Nenhum		

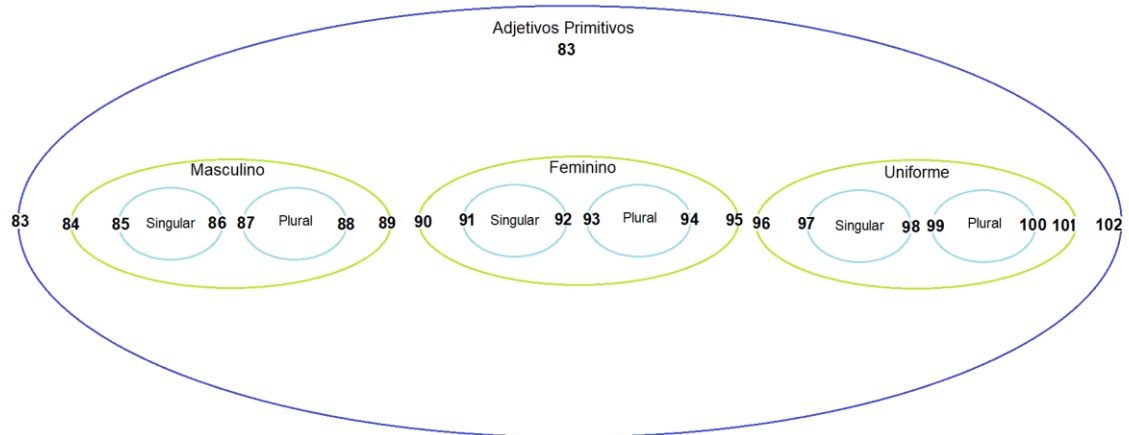
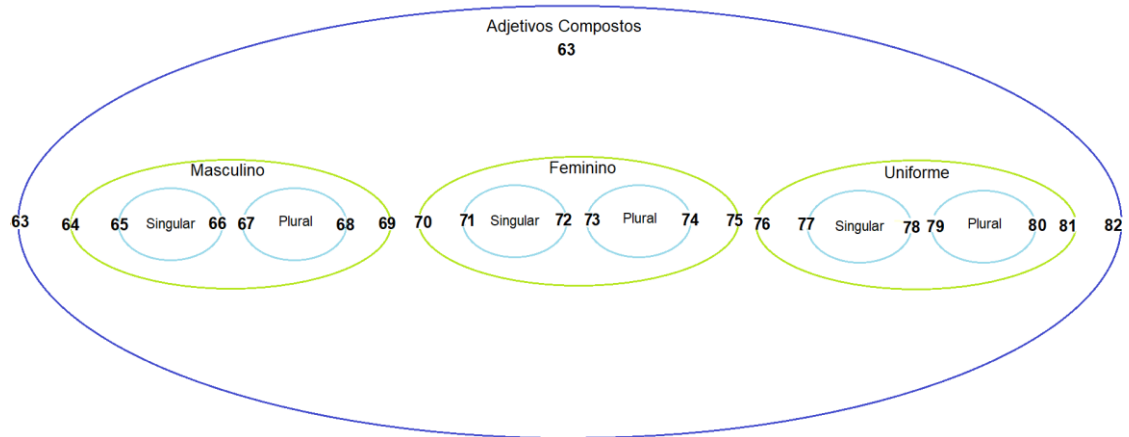
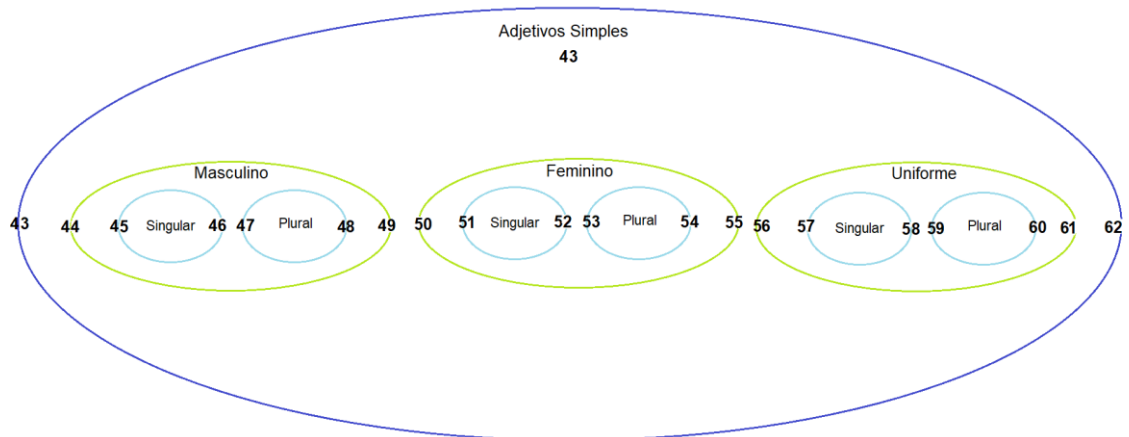
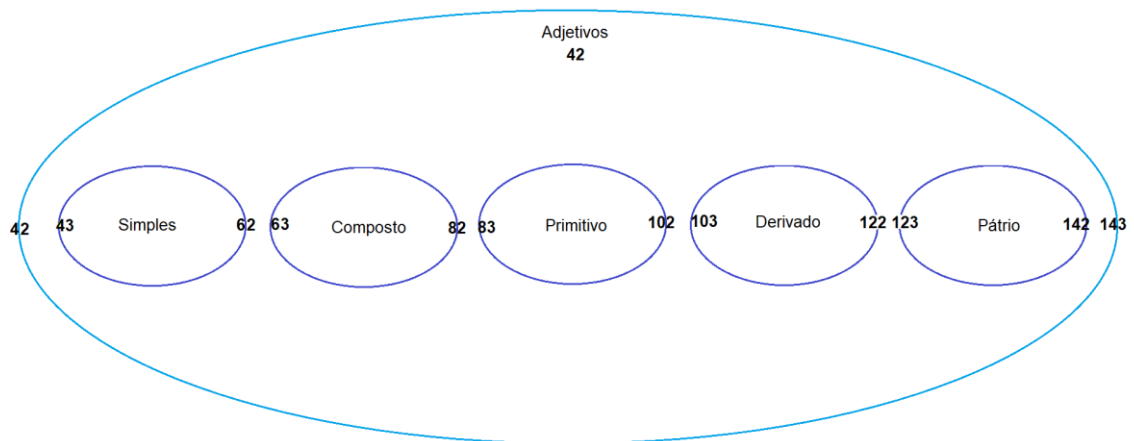
Fig. 3: Estrutura da tabela tokens

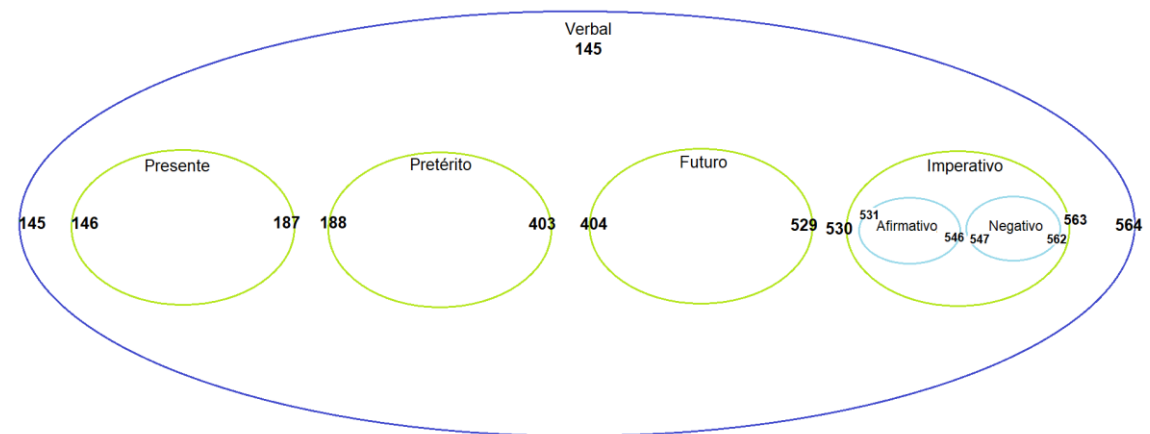
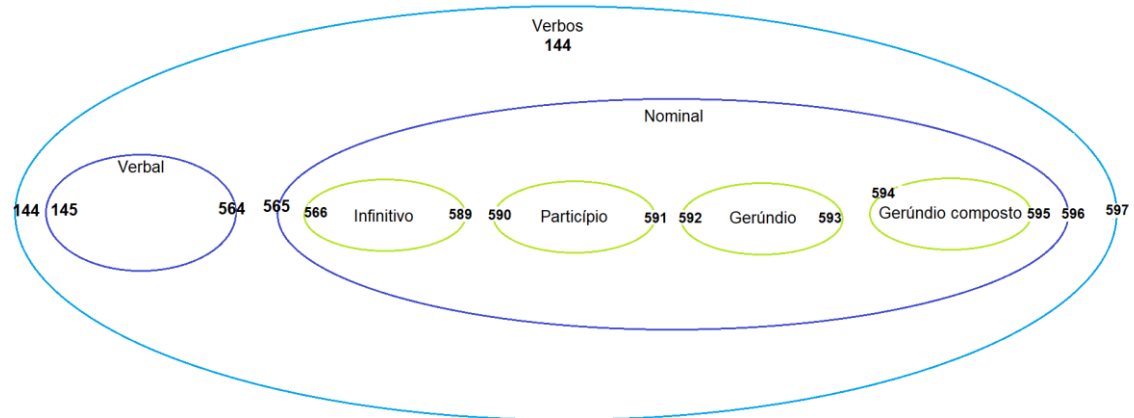
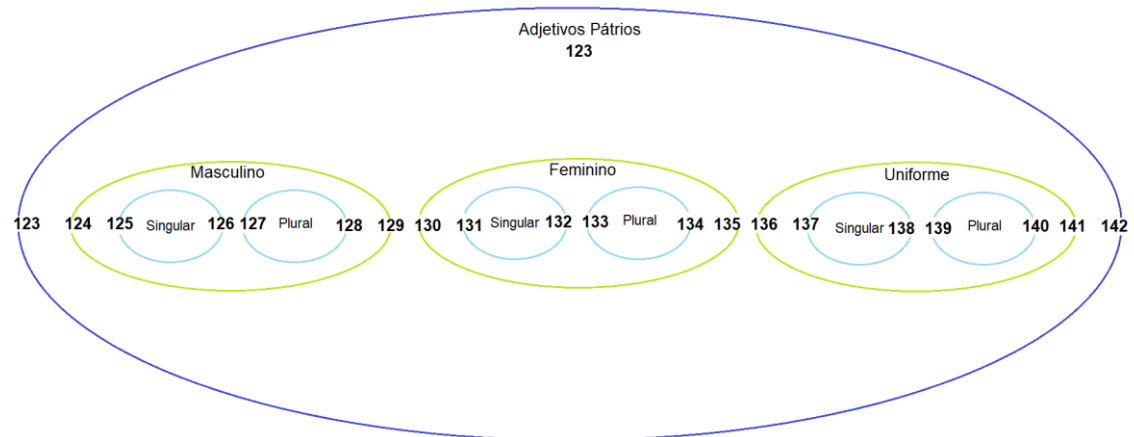
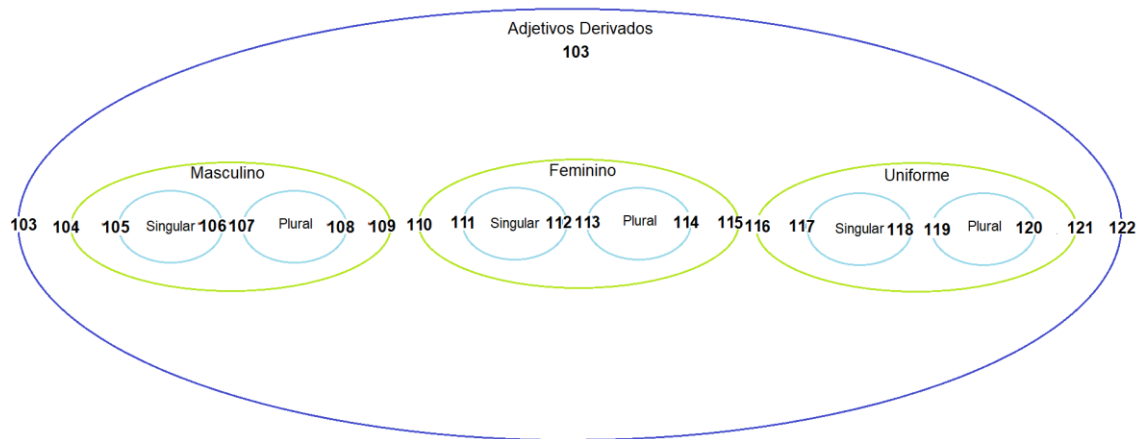
Para definir a estrutura da tabela nested\_classes, é necessário primeiro classificar os tokens de acordo com a hierarquia do tipo árvore mostrada em <https://gitmind.com/app/doc/e408916348>.

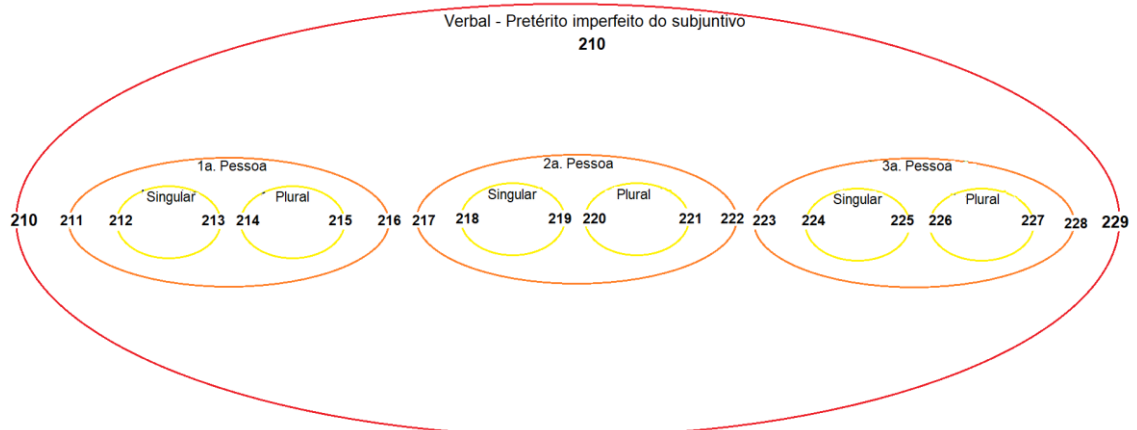
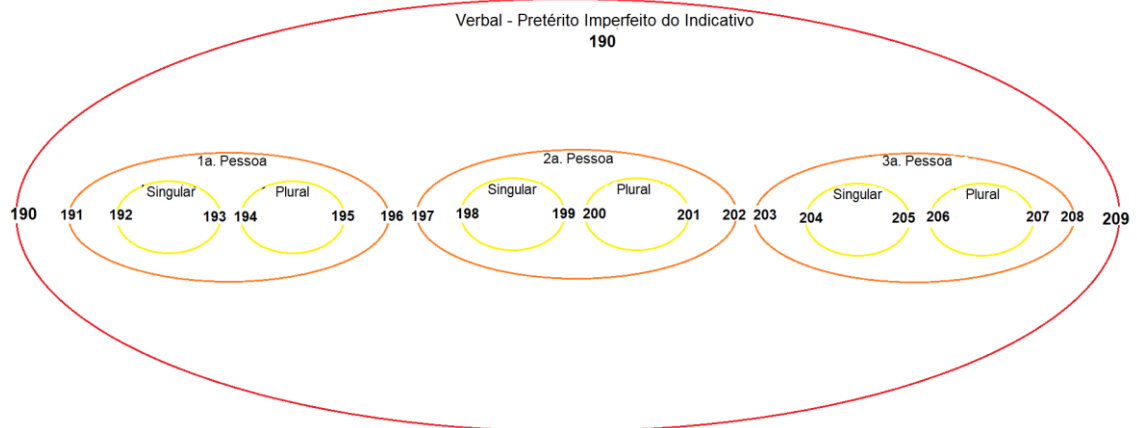
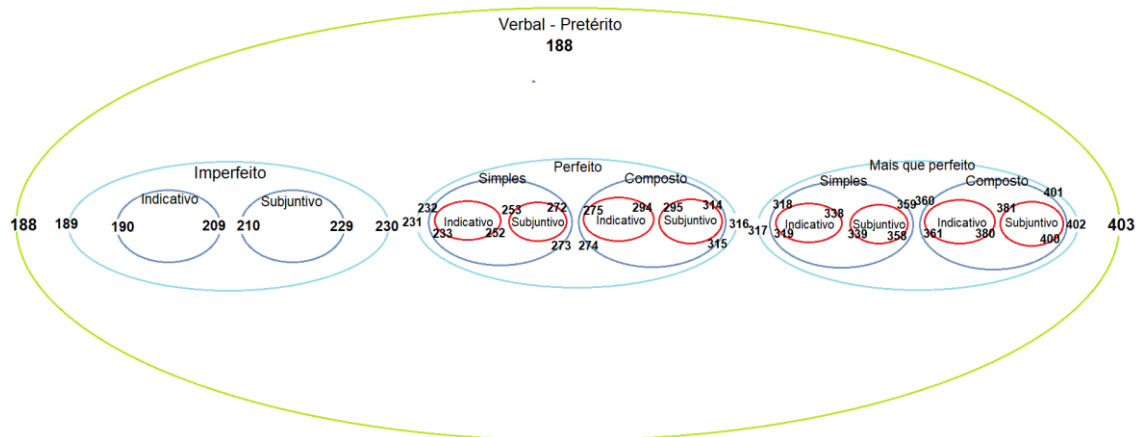
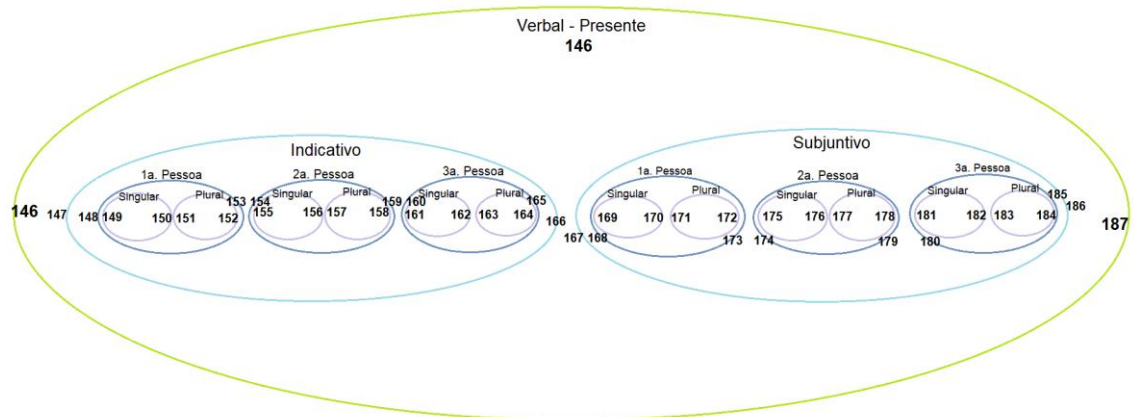
Em seguida, os tokens são agrupados e aninhados de acordo com o “nested set model”, explicado detalhadamente em <http://mikehillier.com/articles/managing-hierarchical-data-in-mysql/>.

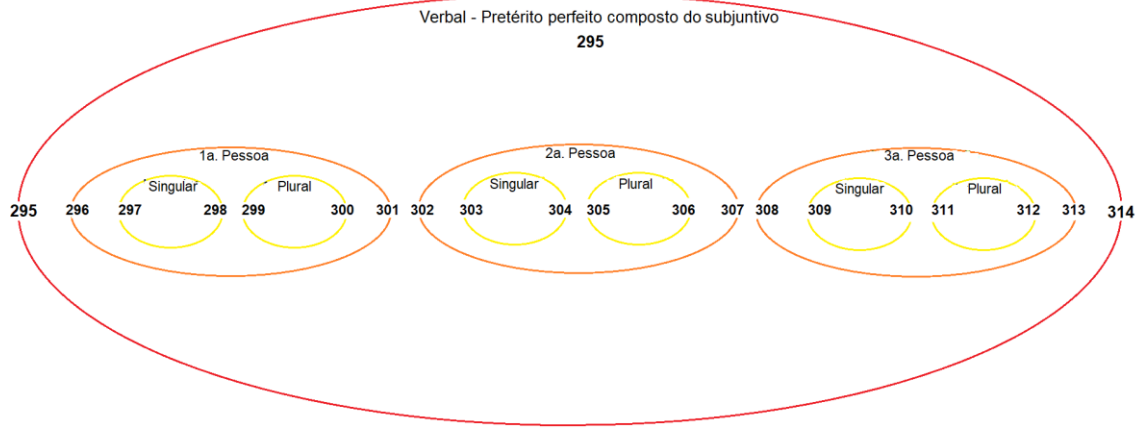
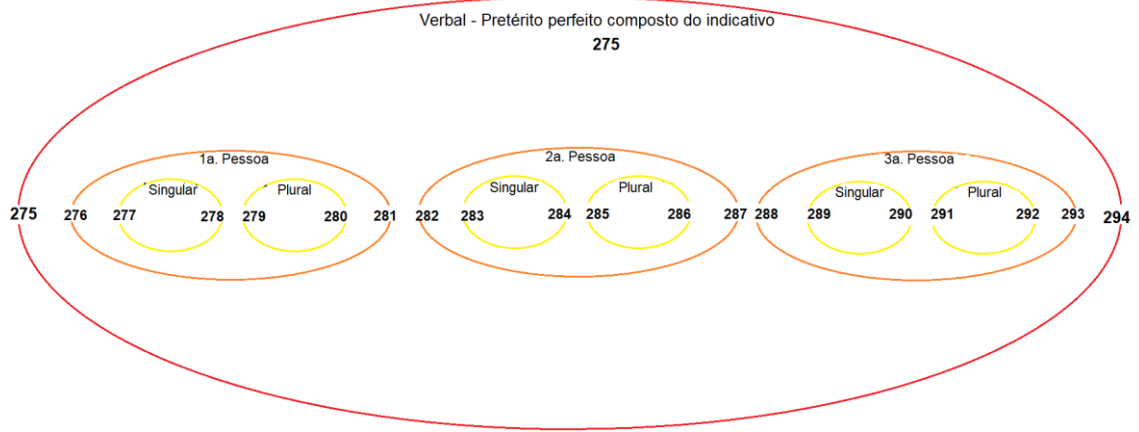
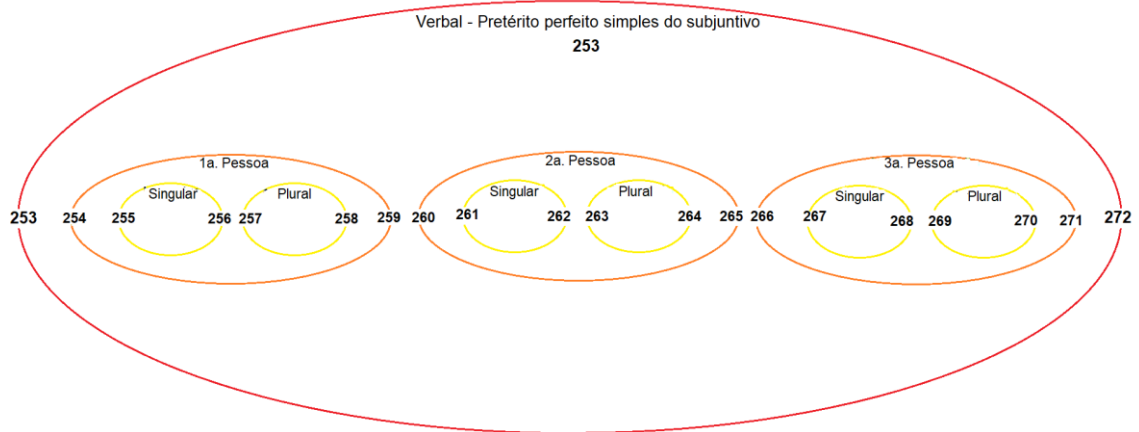
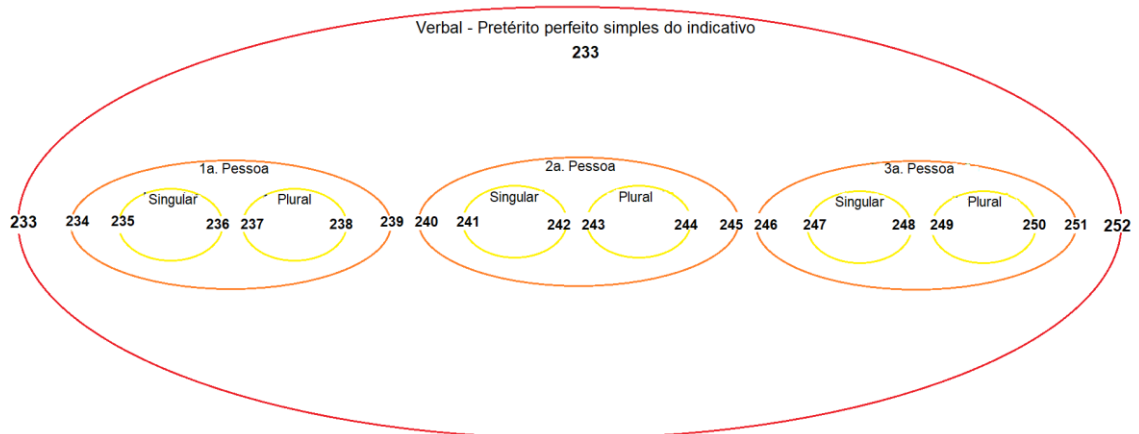
A seguir, temos o resultado da aplicação desse modelo à estrutura gramatical da língua portuguesa:

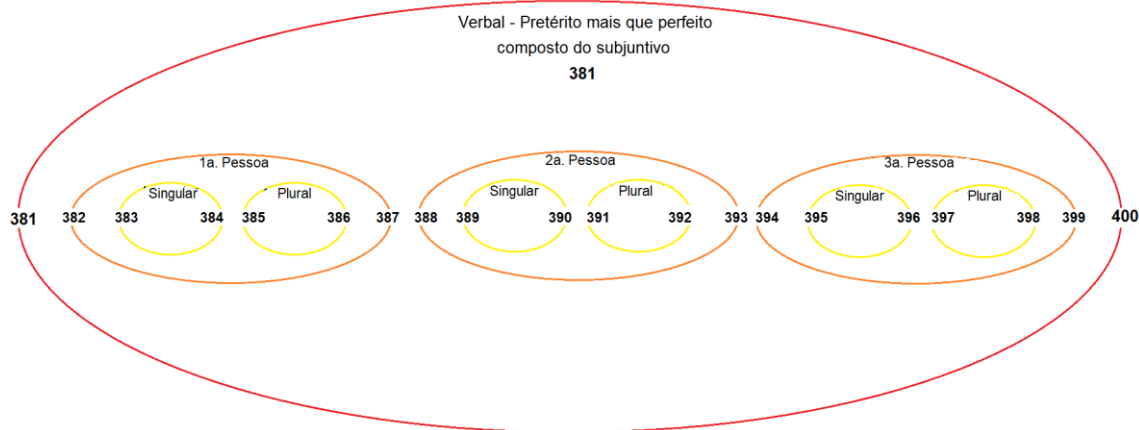
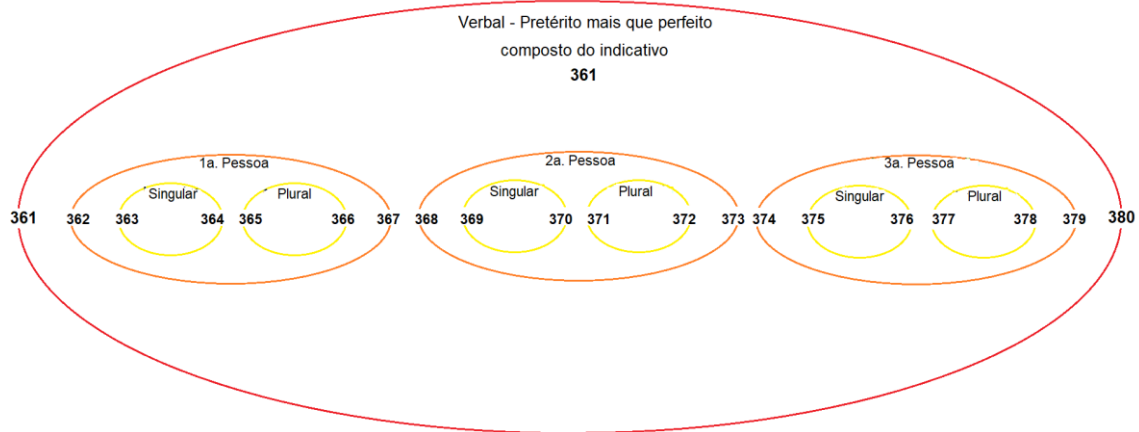
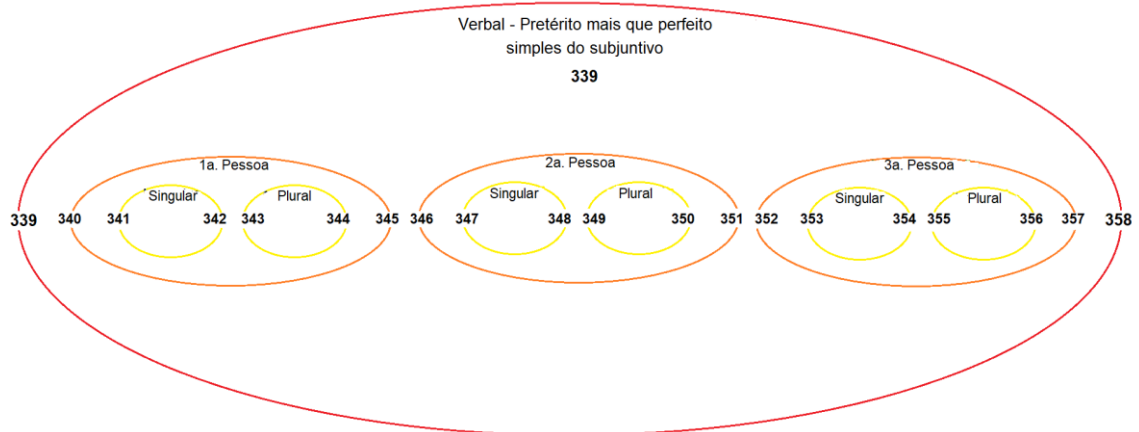
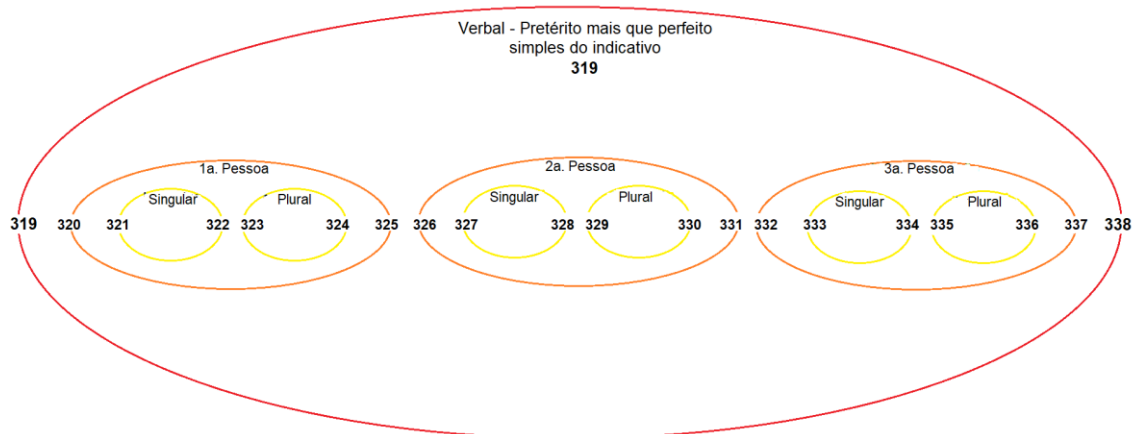


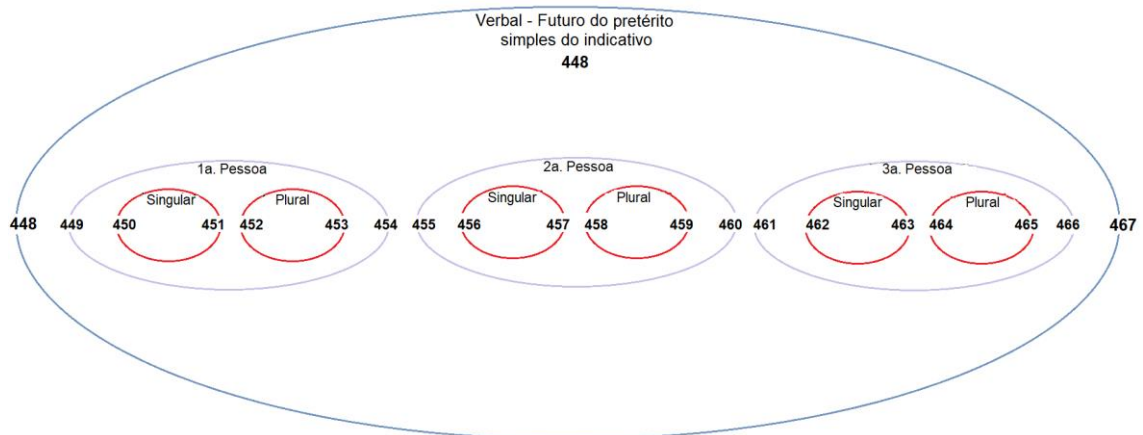
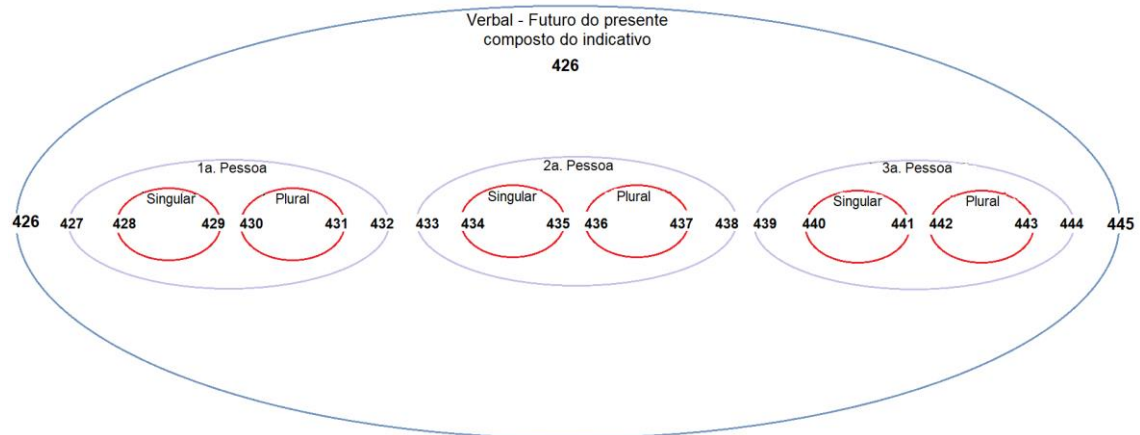
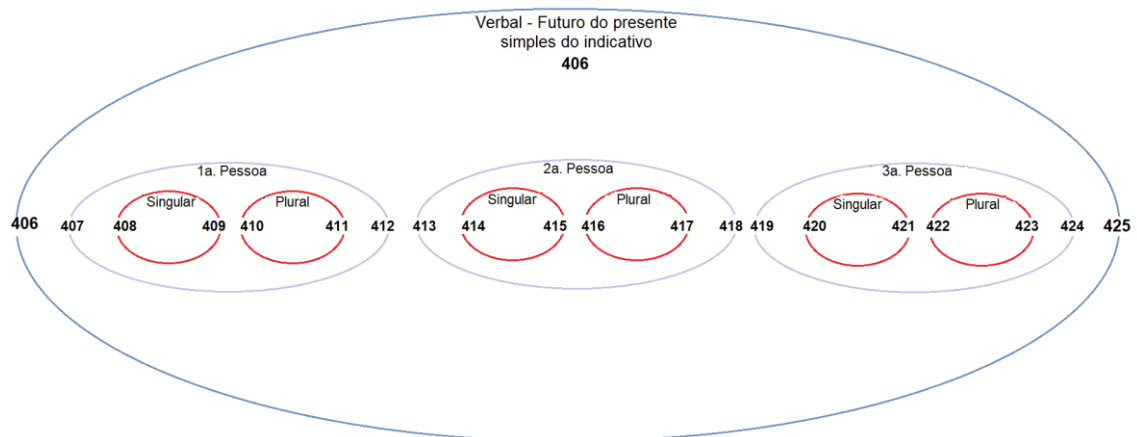
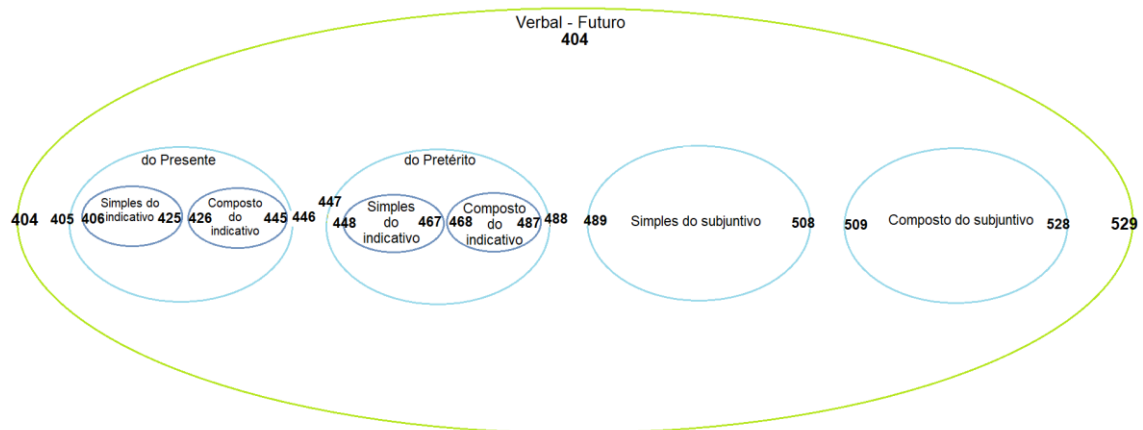




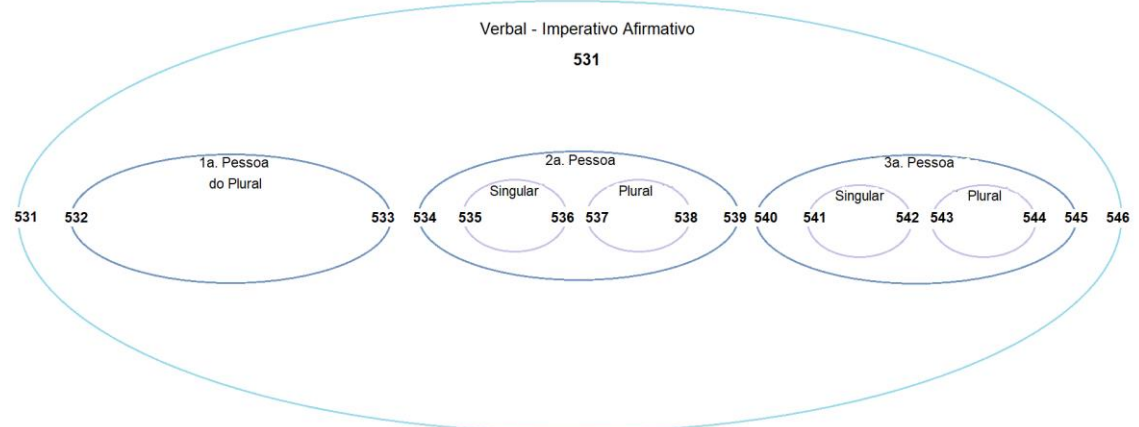
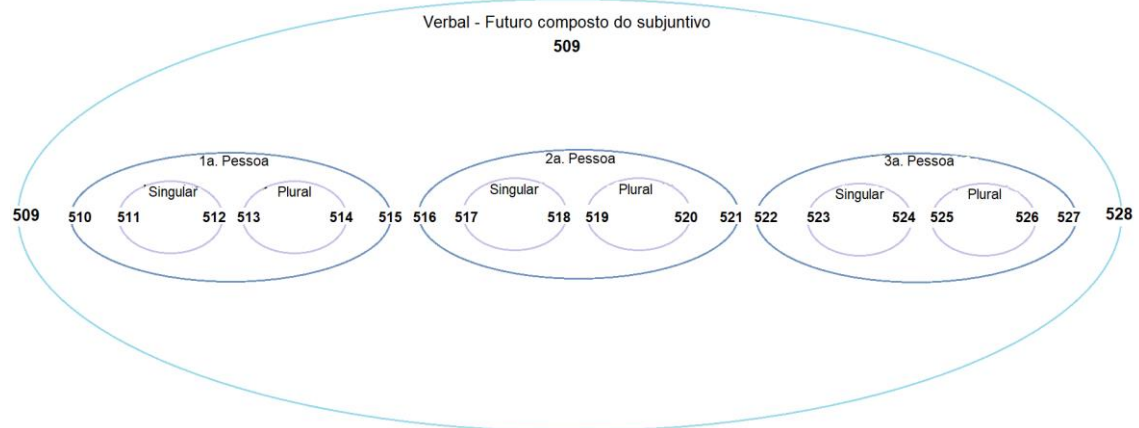
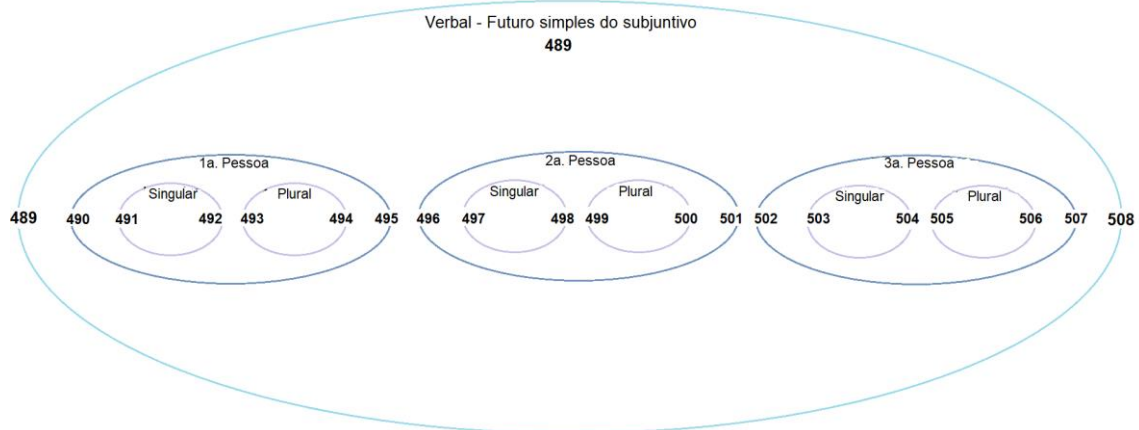
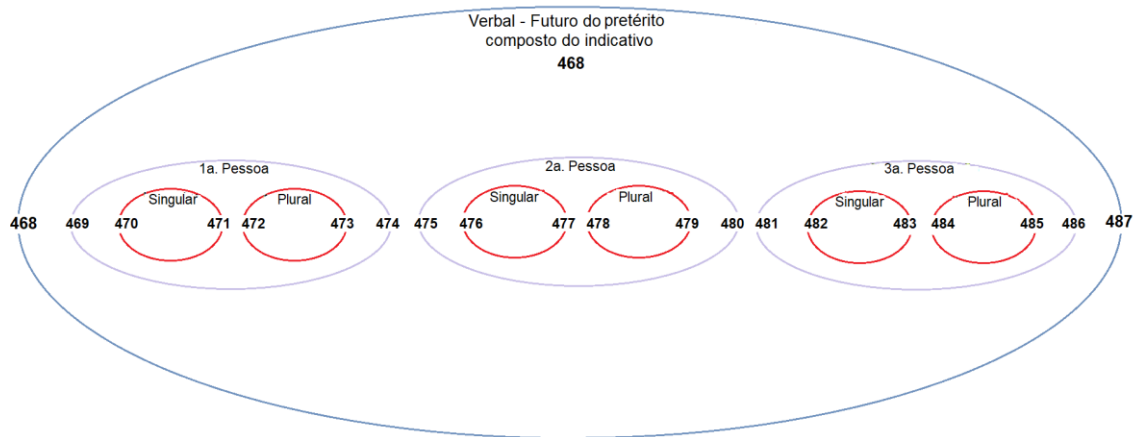


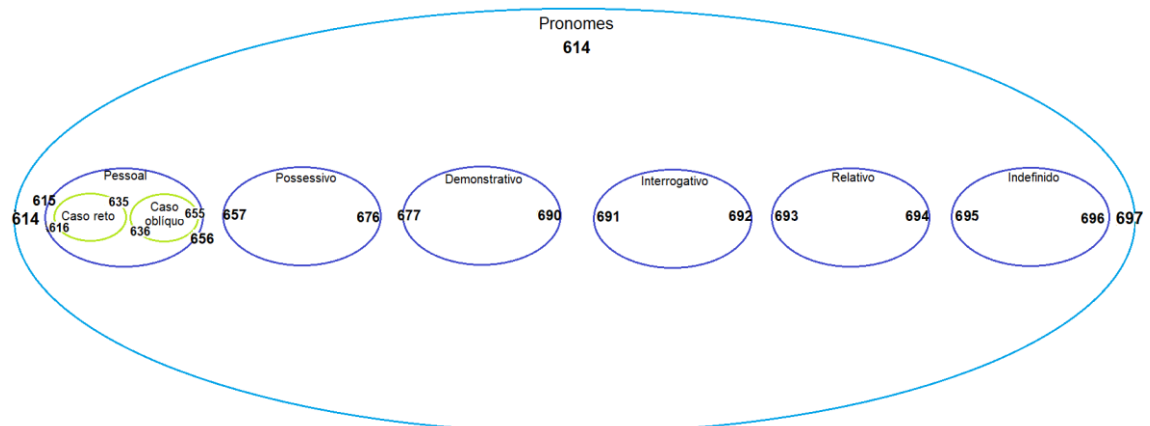
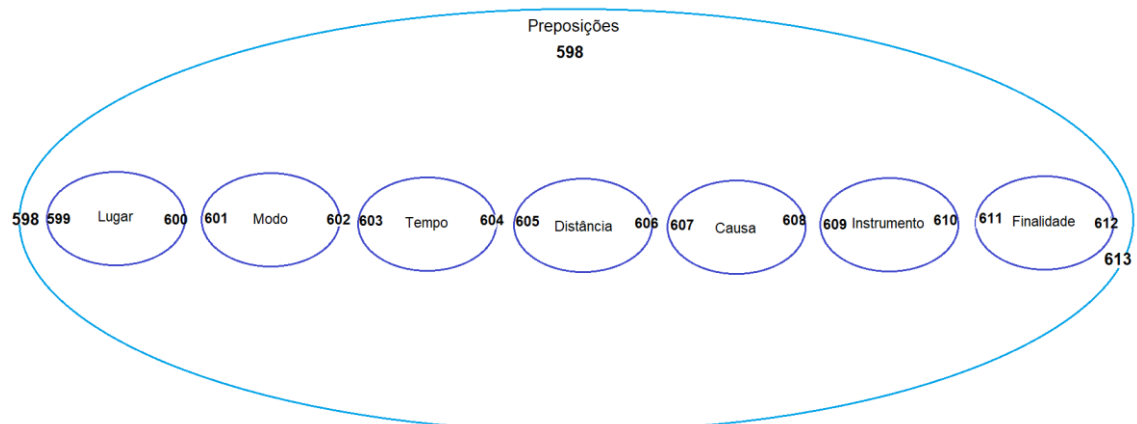
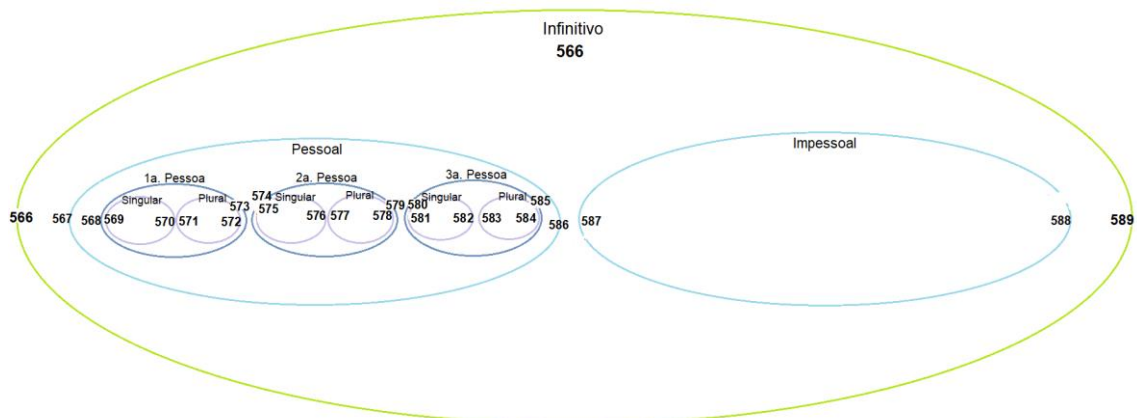
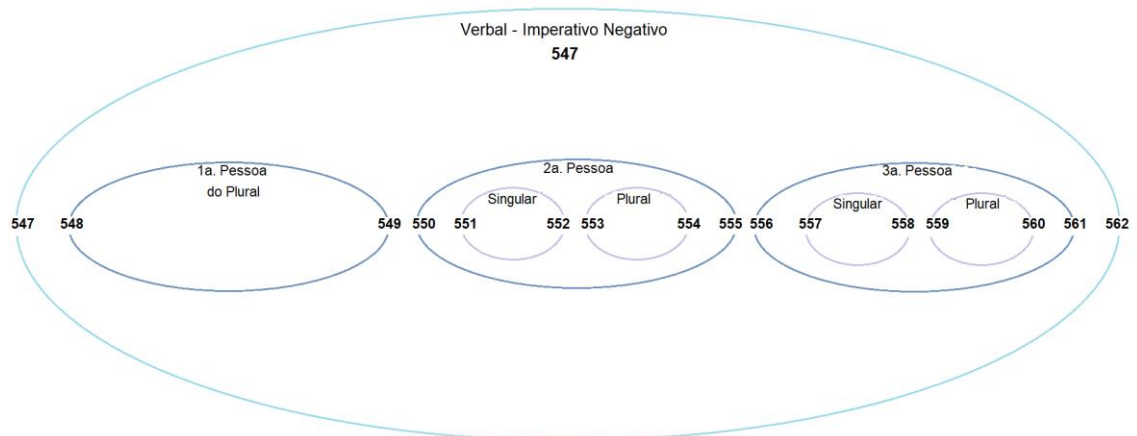


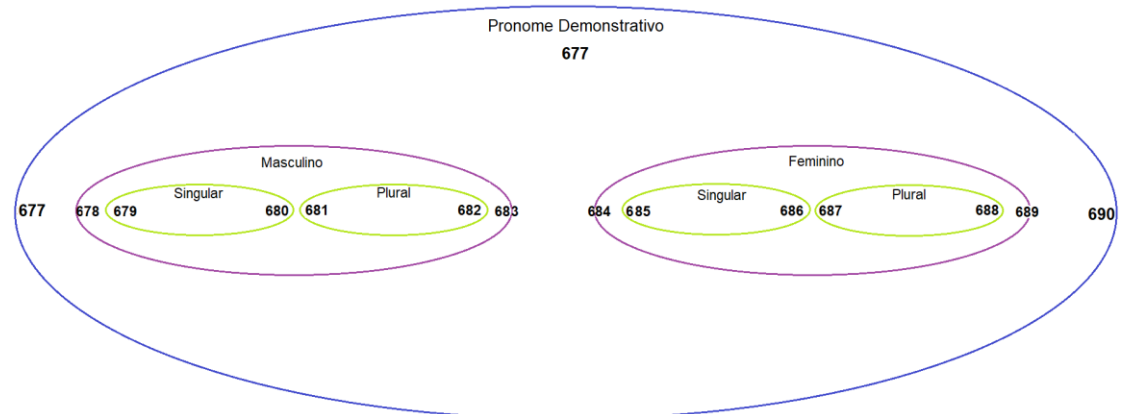
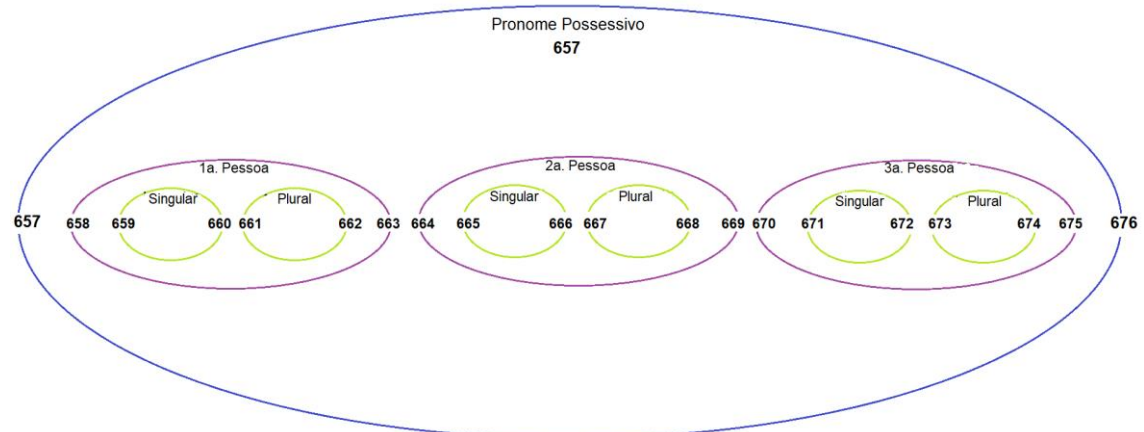
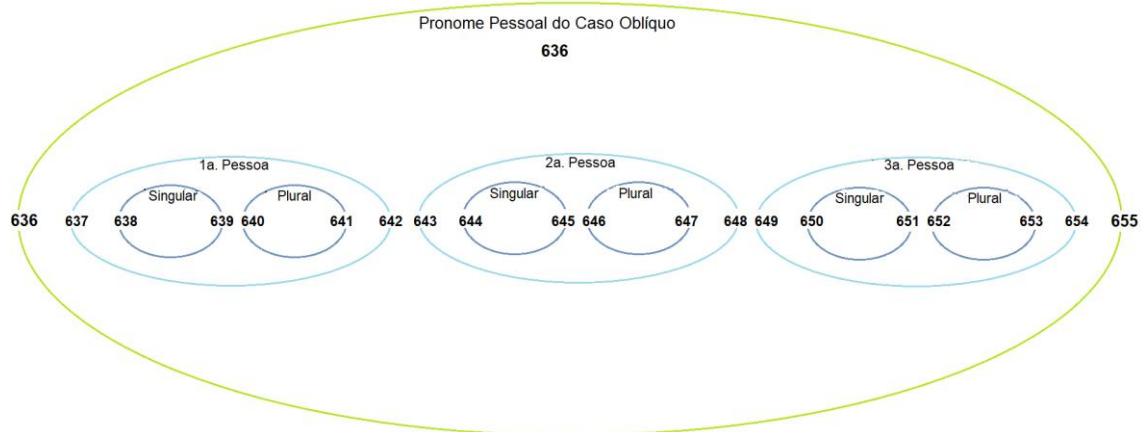
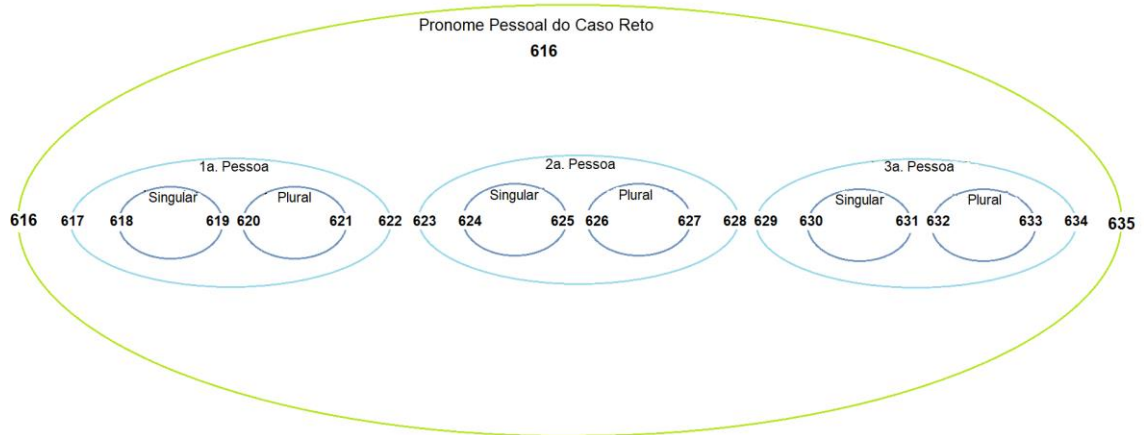


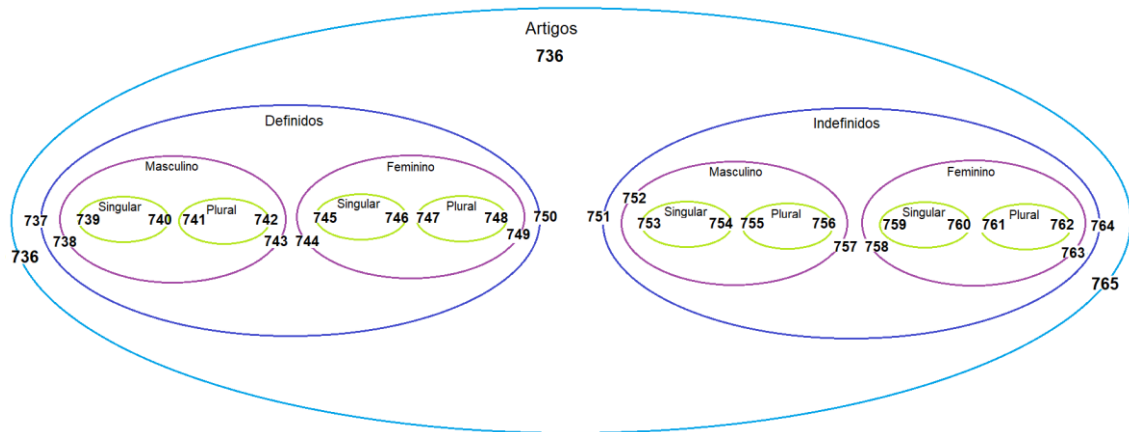
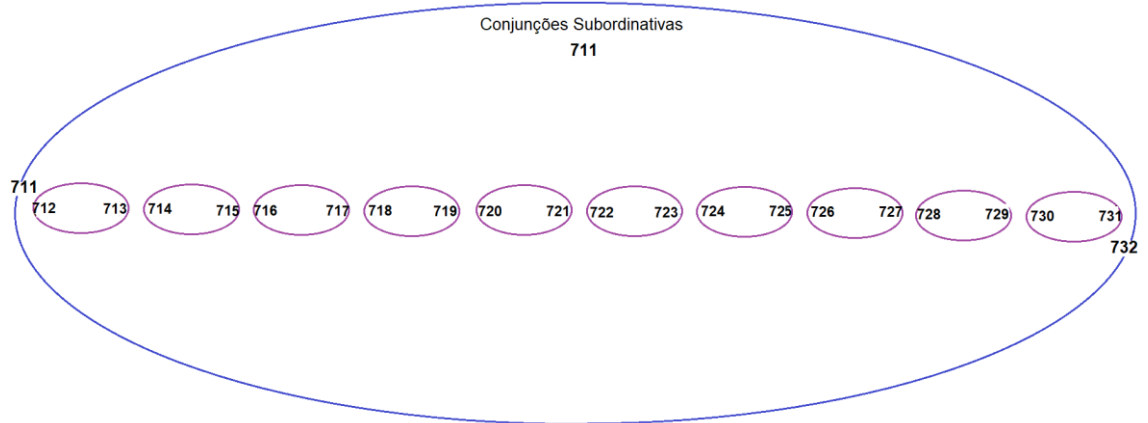
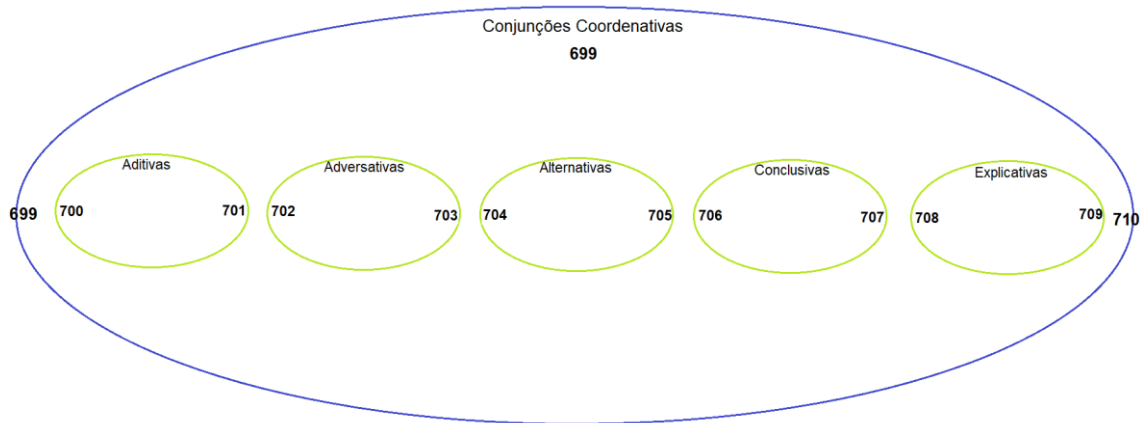
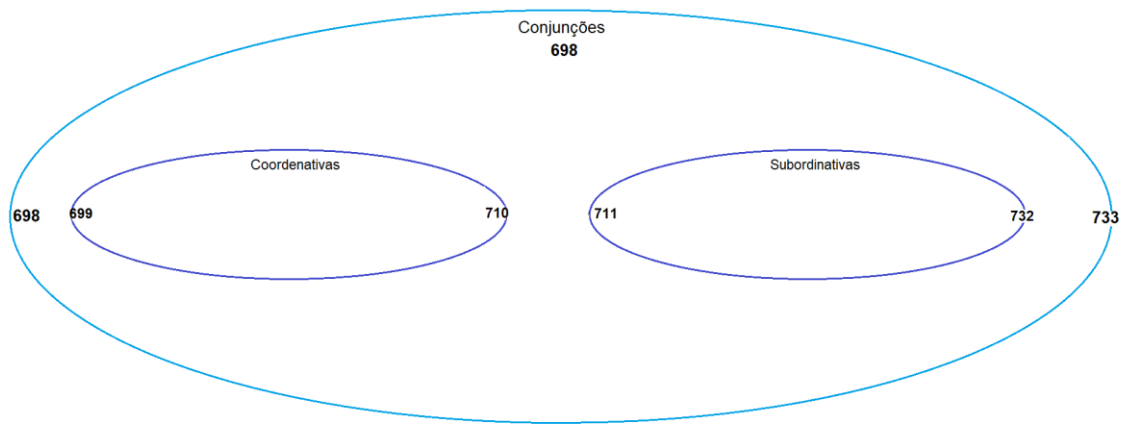














Com todos os conjuntos de classes gramaticais devidamente aninhados e numerados (à direita e à esquerda), o próximo passo é codificar o script SQL para a criação da tabela `nested_classes`. O código utilizado para a criação da tabela é o seguinte:

```
drop table if exists nested_classes;
```

```
drop table if exists nested_classes;
```

```
create TABLE nested_classes (  
    class_id int AUTO_INCREMENT not null,  
    name VARCHAR(200) not null,  
    lft INT not null,  
    rgt INT not null,  
    primary key(class_id));
```

```
INSERT INTO nested_classes values  
    (1, 'TOKEN', 1, 766),  
    (2, 'SUBSTANTIVO', 2, 41),  
    (3, 'COMUM', 3, 34),  
    (4, 'MASCULINO', 4, 9),  
    (5, 'SINGULAR', 5, 6),  
    (6, 'PLURAL', 7, 8),  
    (7, 'FEMININO', 10, 15),  
    (8, 'SINGULAR', 11, 12),  
    (9, 'PLURAL', 13, 14),  
    (10, 'COMUM-DE-DOIS', 16, 21),  
    (11, 'SINGULAR', 17, 18),  
    (12, 'PLURAL', 19, 20),  
    (13, 'EPICENO', 22, 27),  
    (14, 'SINGULAR', 23, 24),  
    (15, 'PLURAL', 25, 26),  
    (16, 'SOBRECOMUM', 28, 33),  
    (17, 'SINGULAR', 29, 30),  
    (18, 'PLURAL', 31, 32),
```

(19, 'PROPRIO', 35, 40),  
(20, 'PESSOA', 36, 37),  
(21, 'CIDADE', 38, 39),  
(22, 'ADJETIVO', 42, 143),  
(23, 'SIMPLES', 43, 62),  
(24, 'MASCULINO', 44, 49),  
(25, 'SINGULAR', 45, 46),  
(26, 'PLURAL', 47, 48),  
(27, 'FEMININO', 50, 55),  
(28, 'SINGULAR', 51, 52),  
(29, 'PLURAL', 53, 54),  
(30, 'UNIFORME', 56, 61),  
(31, 'SINGULAR', 57, 58),  
(32, 'PLURAL', 59, 60),  
(33, 'COMPOSTO', 63, 82),  
(34, 'MASCULINO', 64, 69),  
(35, 'SINGULAR', 65, 66),  
(36, 'PLURAL', 67, 68),  
(37, 'FEMININO', 70, 75),  
(38, 'SINGULAR', 71, 72),  
(39, 'PLURAL', 73, 74),  
(40, 'UNIFORME', 76, 81),  
(41, 'SINGULAR', 77, 78),  
(42, 'PLURAL', 79, 80),  
(43, 'PRIMITIVO', 83, 102),  
(44, 'MASCULINO', 84, 89),  
(45, 'SINGULAR', 85, 86),  
(46, 'PLURAL', 87, 88),  
(47, 'FEMININO', 90, 95),  
(48, 'SINGULAR', 91, 92),  
(49, 'PLURAL', 93, 94),

(50, 'UNIFORME', 96, 101),  
(51, 'SINGULAR', 97, 98),  
(52, 'PLURAL', 99, 100),  
(53, 'DERIVADO', 103, 122),  
(54, 'MASCULINO', 104, 109),  
(55, 'SINGULAR', 105, 106),  
(56, 'PLURAL', 107, 108),  
(57, 'FEMININO', 110, 115),  
(58, 'SINGULAR', 111, 112),  
(59, 'PLURAL', 113, 114),  
(60, 'UNIFORME', 116, 121),  
(61, 'SINGULAR', 117, 118),  
(62, 'PLURAL', 119, 120),  
(63, 'PATRIO', 123, 142),  
(64, 'MASCULINO', 124, 129),  
(65, 'SINGULAR', 125, 126),  
(66, 'PLURAL', 127, 128),  
(67, 'FEMININO', 130, 135),  
(68, 'SINGULAR', 131, 132),  
(69, 'PLURAL', 133, 134),  
(70, 'UNIFORME', 136, 141),  
(71, 'SINGULAR', 137, 138),  
(72, 'PLURAL', 139, 140),  
(73, 'VERBO', 144, 597),  
(74, 'VERBAL', 145, 564),  
(75, 'PRESENTE', 146, 187),  
(76, 'INDICATIVO', 147, 166),  
(77, 'PESSOA 1', 148, 153),  
(78, 'SINGULAR', 149, 150),  
(79, 'PLURAL', 151, 152),  
(80, 'PESSOA 2', 154, 159),



(81, 'SINGULAR', 155, 156),  
(82, 'PLURAL', 157, 158),  
(83, 'PESSOA 3', 160, 165),  
(84, 'SINGULAR', 161, 162),  
(85, 'PLURAL', 163, 164),  
(86, 'SUBJUNTIVO', 167, 186),  
(87, 'PESSOA 1', 168, 173),  
(88, 'SINGULAR', 169, 170),  
(89, 'PLURAL', 171, 172),  
(90, 'PESSOA 2', 174, 179),  
(91, 'SINGULAR', 175, 176),  
(92, 'PLURAL', 177, 178),  
(93, 'PESSOA 3', 180, 185),  
(94, 'SINGULAR', 181, 182),  
(95, 'PLURAL', 183, 184),  
(96, 'PRETERITO', 188, 403),  
(97, 'IMPERFEITO', 189, 230),  
(98, 'INDICATIVO', 190, 209),  
(99, 'PESSOA 1', 191, 196),  
(100, 'SINGULAR', 192, 193),  
(101, 'PLURAL', 194, 195),  
(102, 'PESSOA 2', 197, 202),  
(103, 'SINGULAR', 198, 199),  
(104, 'PLURAL', 200, 201),  
(105, 'PESSOA 3', 203, 208),  
(106, 'SINGULAR', 204, 205),  
(107, 'PLURAL', 206, 207),  
(108, 'SUBJUNTIVO', 210, 229),  
(109, 'PESSOA 1', 211, 216),  
(110, 'SINGULAR', 212, 213),  
(111, 'PLURAL', 214, 215),

(112, 'PESSOA 2', 217, 222),  
(113, 'SINGULAR', 218, 219),  
(114, 'PLURAL', 220, 221),  
(115, 'PESSOA 3', 223, 228),  
(116, 'SINGULAR', 224, 225),  
(117, 'PLURAL', 226, 227),  
(118, 'PERFEITO', 231, 316),  
(119, 'SIMPLES', 232, 273),  
(120, 'INDICATIVO', 233, 252),  
(121, 'PESSOA 1', 234, 239),  
(122, 'SINGULAR', 235, 236),  
(123, 'PLURAL', 237, 238),  
(124, 'PESSOA 2', 240, 245),  
(125, 'SINGULAR', 241, 242),  
(126, 'PLURAL', 243, 244),  
(127, 'PESSOA 3', 246, 251),  
(128, 'SINGULAR', 247, 248),  
(129, 'PLURAL', 249, 250),  
(130, 'SUBJUNTIVO', 253, 272),  
(131, 'PESSOA 1', 254, 259),  
(132, 'SINGULAR', 255, 256),  
(133, 'PLURAL', 257, 258),  
(134, 'PESSOA 2', 260, 265),  
(135, 'SINGULAR', 261, 262),  
(136, 'PLURAL', 263, 264),  
(137, 'PESSOA 3', 266, 271),  
(138, 'SINGULAR', 267, 268),  
(139, 'PLURAL', 269, 270),  
(140, 'COMPOSTO', 274, 315),  
(141, 'INDICATIVO', 275, 294),  
(142, 'PESSOA 1', 276, 281),

(143, 'SINGULAR', 277, 278),  
(144, 'PLURAL', 279, 280),  
(145, 'PESSOA 2', 282, 287),  
(146, 'SINGULAR', 283, 284),  
(147, 'PLURAL', 285, 286),  
(148, 'PESSOA 3', 288, 293),  
(149, 'SINGULAR', 289, 290),  
(150, 'PLURAL', 291, 292),  
(151, 'SUBJUNTIVO', 295, 314),  
(152, 'PESSOA 1', 296, 301),  
(153, 'SINGULAR', 297, 298),  
(154, 'PLURAL', 299, 300),  
(155, 'PESSOA 2', 302, 307),  
(156, 'SINGULAR', 303, 304),  
(157, 'PLURAL', 305, 306),  
(158, 'PESSOA 3', 308, 313),  
(159, 'SINGULAR', 309, 310),  
(160, 'PLURAL', 311, 312),  
(161, 'MAIS QUE PERFEITO', 317, 402),  
(162, 'SIMPLES', 318, 359),  
(163, 'INDICATIVO', 319, 338),  
(164, 'PESSOA 1', 320, 325),  
(165, 'SINGULAR', 321, 322),  
(166, 'PLURAL', 323, 324),  
(167, 'PESSOA 2', 326, 331),  
(168, 'SINGULAR', 327, 328),  
(169, 'PLURAL', 329, 330),  
(170, 'PESSOA 3', 332, 337),  
(171, 'SINGULAR', 333, 334),  
(172, 'PLURAL', 335, 336),  
(173, 'SUBJUNTIVO', 339, 358),

(174, 'PESSOA 1', 340, 345),  
(175, 'SINGULAR', 341, 342),  
(176, 'PLURAL', 343, 344),  
(177, 'PESSOA 2', 346, 351),  
(178, 'SINGULAR', 347, 348),  
(179, 'PLURAL', 349, 350),  
(180, 'PESSOA 3', 352, 357),  
(181, 'SINGULAR', 353, 354),  
(182, 'PLURAL', 355, 356),  
(183, 'COMPOSTO', 360, 401),  
(184, 'INDICATIVO', 361, 380),  
(185, 'PESSOA 1', 362, 367),  
(186, 'SINGULAR', 363, 364),  
(187, 'PLURAL', 365, 366),  
(188, 'PESSOA 2', 368, 373),  
(189, 'SINGULAR', 369, 370),  
(190, 'PLURAL', 371, 372),  
(191, 'PESSOA 3', 374, 379),  
(192, 'SINGULAR', 375, 376),  
(193, 'PLURAL', 377, 378),  
(194, 'SUBJUNTIVO', 381, 400),  
(195, 'PESSOA 1', 382, 387),  
(196, 'SINGULAR', 383, 384),  
(197, 'PLURAL', 385, 386),  
(198, 'PESSOA 2', 388, 393),  
(199, 'SINGULAR', 389, 390),  
(200, 'PLURAL', 391, 392),  
(201, 'PESSOA 3', 394, 399),  
(202, 'SINGULAR', 395, 396),  
(203, 'PLURAL', 397, 398),  
(204, 'FUTURO', 404, 529),

(205, 'DO PRESENTE', 405, 446),  
(206, 'SIMPLES DO INDICATIVO', 406, 425),  
(207, 'PESSOA 1', 407, 412),  
(208, 'SINGULAR', 408, 409),  
(209, 'PLURAL', 410, 411),  
(210, 'PESSOA 2', 413, 418),  
(211, 'SINGULAR', 414, 415),  
(212, 'PLURAL', 416, 417),  
(213, 'PESSOA 3', 419, 424),  
(214, 'SINGULAR', 420, 421),  
(215, 'PLURAL', 422, 423),  
(216, 'COMPOSTO DO INDICATIVO', 426, 445),  
(217, 'PESSOA 1', 427, 432),  
(218, 'SINGULAR', 428, 429),  
(219, 'PLURAL', 430, 431),  
(220, 'PESSOA 2', 433, 438),  
(221, 'SINGULAR', 434, 435),  
(222, 'PLURAL', 436, 437),  
(223, 'PESSOA 3', 439, 444),  
(224, 'SINGULAR', 440, 441),  
(225, 'PLURAL', 442, 443),  
(226, 'DO PRETERITO', 447, 488),  
(227, 'SIMPLES DO INDICATIVO', 448, 467),  
(228, 'PESSOA 1', 449, 454),  
(229, 'SINGULAR', 450, 451),  
(230, 'PLURAL', 452, 453),  
(231, 'PESSOA 2', 455, 459),  
(232, 'SINGULAR', 456, 457),  
(233, 'PLURAL', 458, 459),  
(234, 'PESSOA 3', 461, 466),  
(235, 'SINGULAR', 462, 463),

(236, 'PLURAL', 464, 465),  
(237, 'COMPOSTO DO INDICATIVO', 468, 487),  
(238, 'PESSOA 1', 469, 474),  
(239, 'SINGULAR', 470, 471),  
(240, 'PLURAL', 472, 473),  
(241, 'PESSOA 2', 475, 480),  
(242, 'SINGULAR', 476, 477),  
(243, 'PLURAL', 478, 479),  
(244, 'PESSOA 3', 481, 486),  
(245, 'SINGULAR', 482, 483),  
(246, 'PLURAL', 484, 485),  
(247, 'SIMPLES DO SUBJUNTIVO', 489, 508),  
(248, 'PESSOA 1', 490, 495),  
(249, 'SINGULAR', 491, 492),  
(250, 'PLURAL', 493, 494),  
(251, 'PESSOA 2', 496, 501),  
(252, 'SINGULAR', 497, 498),  
(253, 'PLURAL', 499, 500),  
(254, 'PESSOA 3', 502, 507),  
(255, 'SINGULAR', 503, 504),  
(256, 'PLURAL', 505, 506),  
(257, 'COMPOSTO DO SUBJUNTIVO', 509, 528),  
(258, 'PESSOA 1', 510, 515),  
(259, 'SINGULAR', 511, 512),  
(260, 'PLURAL', 513, 514),  
(261, 'PESSOA 2', 516, 521),  
(262, 'SINGULAR', 517, 518),  
(263, 'PLURAL', 519, 520),  
(264, 'PESSOA 3', 522, 527),  
(265, 'SINGULAR', 523, 524),  
(266, 'PLURAL', 525, 526),

(267, 'IMPERATIVO', 530, 563),  
(268, 'AFIRMATIVO', 531, 546),  
(269, 'PESSOA 1 PLURAL', 532, 533),  
(270, 'PESSOA 2', 534, 539),  
(271, 'SINGULAR', 535, 536),  
(272, 'PLURAL', 537, 538),  
(273, 'PESSOA 3', 540, 545),  
(274, 'SINGULAR', 541, 542),  
(275, 'PLURAL', 543, 544),  
(276, 'NEGATIVO', 547, 562),  
(277, 'PESSOA 1 PLURAL', 548, 549),  
(278, 'PESSOA 2', 550, 555),  
(279, 'SINGULAR', 551, 552),  
(280, 'PLURAL', 553, 554),  
(281, 'PESSOA 3', 556, 561),  
(282, 'SINGULAR', 557, 558),  
(283, 'PLURAL', 559, 560),  
(284, 'NOMINAL', 565, 596),  
(285, 'INFINITIVO', 566, 589),  
(286, 'PESSOAL', 567, 586),  
(287, 'PESSOA 1', 568, 573),  
(288, 'SINGULAR', 569, 570),  
(289, 'PLURAL', 571, 572),  
(290, 'PESSOA 2', 574, 579),  
(291, 'SINGULAR', 575, 576),  
(292, 'PLURAL', 577, 578),  
(293, 'PESSOA 3', 580, 585),  
(294, 'SINGULAR', 581, 582),  
(295, 'PLURAL', 583, 584),  
(296, 'IMPESSOAL', 587, 588),  
(297, 'PARTICÍPIO', 590, 591),

(298, 'GERUNDIO', 592, 593),  
(299, 'GERUNDIO COMPOSTO', 594, 595),  
(300, 'PREPOSICOES', 598, 613),  
(301, 'DE LUGAR', 599, 600),  
(302, 'DE MODO', 601, 602),  
(303, 'DE TEMPO', 603, 604),  
(304, 'DE DISTANCIA', 605, 606),  
(305, 'DE CAUSA', 607, 608),  
(306, 'DE INSTRUMENTO', 609, 610),  
(307, 'DE FINALIDADE', 611, 612),  
(308, 'PRONOME', 614, 697),  
(309, 'PESSOAL', 615, 656),  
(310, 'CASO RETO', 616, 635),  
(311, 'PESSOA 1', 617, 622),  
(312, 'SINGULAR', 618, 619),  
(313, 'PLURAL', 620, 621),  
(314, 'PESSOA 2', 623, 628),  
(315, 'SINGULAR', 624, 625),  
(316, 'PLURAL', 626, 627),  
(317, 'PESSOA 3', 629, 634),  
(318, 'SINGULAR', 630, 631),  
(319, 'PLURAL', 632, 633),  
(320, 'CASO OBLIQUO', 636, 655),  
(321, 'PESSOA 1', 637, 642),  
(322, 'SINGULAR', 638, 639),  
(323, 'PLURAL', 640, 641),  
(324, 'PESSOA 2', 643, 648),  
(325, 'SINGULAR', 644, 645),  
(326, 'PLURAL', 646, 647),  
(327, 'PESSOA 3', 649, 654),  
(328, 'SINGULAR', 650, 651),



(329, 'PLURAL', 652, 653),  
(330, 'POSSESSIVO', 657, 676),  
(331, 'PESSOA 1', 658, 663),  
(332, 'SINGULAR', 659, 660),  
(333, 'PLURAL', 661, 662),  
(334, 'PESSOA 2', 664, 669),  
(335, 'SINGULAR', 665, 666),  
(336, 'PLURAL', 667, 668),  
(337, 'PESSOA 3', 670, 675),  
(338, 'SINGULAR', 671, 672),  
(339, 'PLURAL', 673, 674),  
(340, 'DEMONSTRATIVO', 677, 690),  
(341, 'MASCULINO', 678, 683),  
(342, 'SINGULAR', 679, 680),  
(343, 'PLURAL', 681, 682),  
(344, 'FEMININO', 684, 689),  
(345, 'SINGULAR', 685, 686),  
(346, 'PLURAL', 687, 688),  
(347, 'INTERROGATIVO', 691, 692),  
(348, 'RELATIVO', 693, 694),  
(349, 'INDEFINIDO', 695, 696),  
(350, 'CONJUNCAO', 698, 733),  
(351, 'COORDENATIVA', 699, 710),  
(352, 'ADITIVA', 700, 701),  
(353, 'ADVERSATIVA', 702, 703),  
(354, 'ALTERNATIVA', 704, 705),  
(355, 'CONCLUSIVA', 706, 707),  
(356, 'EXPLICATIVA', 708, 709),  
(357, 'SUBORDINATIVA', 711, 732),  
(358, 'CAUSAL', 712, 713),  
(359, 'CONCESSIVA', 714, 715),

(360, 'CONDICIONAL', 716, 717),  
(361, 'CONFORMATIVA', 718, 719),  
(362, 'FINAL', 720, 721),  
(363, 'PROPORCIONAL', 722, 723),  
(364, 'TEMPORAL', 724, 725),  
(365, 'COMPARATIVA', 726, 727),  
(366, 'CONSECUTIVA', 728, 729),  
(367, 'INTEGRANTE', 730, 731),  
(368, 'PONTUACAO', 734, 735),  
(369, 'ARTIGO', 736, 765),  
(370, 'DEFINIDO', 737, 750),  
(371, 'MASCULINO', 738, 743),  
(372, 'SINGULAR', 739, 740),  
(373, 'PLURAL', 741, 742),  
(374, 'FEMININO', 744, 749),  
(375, 'SINGULAR', 745, 746),  
(376, 'PLURAL', 747, 748),  
(377, 'INDEFINIDO', 751, 764),  
(378, 'MASCULINO', 752, 757),  
(379, 'SINGULAR', 753, 754),  
(380, 'PLURAL', 755, 756),  
(381, 'FEMININO', 758, 763),  
(382, 'SINGULAR', 759, 760),  
(383, 'PLURAL', 761, 762);

Para armazenar os tokens, precisamos de uma segunda tabela, que recebeu o nome de “tokens”. Para criar essa tabela, foi utilizado o seguinte script:

```
drop table if exists tokens;
```

```
CREATE TABLE tokens
```

```
(
```

```
    token_id INT AUTO_INCREMENT PRIMARY KEY,
```

```
    name VARCHAR(200),
```

```
    class_id INT NOT NULL
```

```
);
```

A maneira mais eficiente de popular a tabela tokens é através de uma interface que permita economia de digitação e automatização de processos.

Sendo assim, a interface deve permitir a importação de lista de tokens, e também a geração de tokens a partir de um elemento comum, como é o caso dos verbos. Além disso, o registro manual de tokens deve ser facilitado através de recursos que auxiliem na identificação de sua classificação e numeração dentro do diagrama do “nested set model”.