

---

# Spis treści

<b>Przedmowa</b>	<b>13</b>
<b>Informacja o finansowaniu prac</b>	<b>15</b>
<b>Podziękowania</b>	<b>17</b>
<b>1. Założenia badawcze</b>	<b>19</b>
1.1. Referencja, koreferencja, anafora, asocjacja . . . . .	19
1.2. Motywacja . . . . .	21
1.3. Cele badawcze . . . . .	22
1.4. Zakres badań . . . . .	23
1.5. Metodologia . . . . .	24
<b>2. Od ujęć teoretycznych do dekodowania relacji referencyjnych</b>	<b>27</b>
2.1. Pojęcie i zakres referencji . . . . .	27
2.2. Klasyfikacje typów wzmianek i relacji referencyjnych . . . . .	29
2.2.1. Klemensiewicz . . . . .	30
2.2.2. Topolińska . . . . .	31
2.2.3. Paduczewa . . . . .	32
2.2.4. Clark i inne klasyfikacje zagraniczne . . . . .	33
2.3. Cechy relacji referencyjnych . . . . .	35
2.4. Projekty korpusowe . . . . .	38
2.5. Komputerowe implementacje modelu referencji . . . . .	41
2.6. Metody ewaluacji . . . . .	48
2.6.1. Miara MUC . . . . .	50
2.6.2. Miara B <sup>3</sup> . . . . .	51
2.6.3. Miara CEAF . . . . .	52
2.6.4. Miara BLANC . . . . .	53
<b>3. Model relacji referencyjnych</b>	<b>55</b>
3.1. Świat tekstu i własność referencji . . . . .	55
3.2. Typy i granice wzmianek . . . . .	56

---

3.3. Relacje tekstowe i pozatekstowe . . . . .	58
3.4. Typologia referencyjna . . . . .	59
3.4.1. Koreferencja . . . . .	61
3.4.2. Referencja pośrednia . . . . .	61
3.4.3. Relacje wspierające . . . . .	63
3.4.4. Relacje wykluczające . . . . .	63
3.4.5. Aspekty . . . . .	64
<b>4. Korpus zależności referencyjnych . . . . .</b>	<b>67</b>
4.1. Wybór tekstów . . . . .	67
4.2. Wybór strategii anotacyjnej . . . . .	69
4.2.1. Liczba i profil anotatorów . . . . .	69
4.2.2. Anotacja szeregową a anotacja równoległą . . . . .	71
4.2.3. Preatotacja . . . . .	72
4.2.4. Superanotacja automatyczna . . . . .	73
4.3. Prace anotacyjne . . . . .	75
4.3.1. Faza rozpoznawcza . . . . .	75
4.3.2. Anotacja koreferencji nominalnej . . . . .	76
4.3.3. Anotacja ogólnych zależności referencyjnych . . . . .	78
4.4. Narzędzia anotacyjne . . . . .	79
4.5. Zgodność anotatorów . . . . .	83
4.5.1. Wzmianki . . . . .	83
4.5.2. Klastry koreferencyjne . . . . .	84
4.5.3. Pozostałe relacje . . . . .	85
4.6. Korekta błędów . . . . .	87
4.7. Udostępnienie korpusu . . . . .	88
4.7.1. Format SemEval/CoNLL . . . . .	89
4.7.2. Format MMAX . . . . .	91
4.7.3. Format TEI . . . . .	94
4.7.4. Format narzędzia BRAT i wersja online korpusu . . . . .	96
4.7.5. Wyszukiwarka korpusowa . . . . .	101
4.8. Statystyki korpusowe . . . . .	101
4.8.1. Własności tekstów . . . . .	101
4.8.2. Własności wzmianek . . . . .	103
4.8.3. Statystyka relacji referencyjnych . . . . .	107

---

<b>5. Implementacja</b>	<b>113</b>
5.1. Wykrywanie wzmianek . . . . .	113
5.1.1. System regułowy . . . . .	114
5.1.2. System statystyczny . . . . .	115
5.2. Wykrywanie koreferencji . . . . .	117
5.2.1. System regułowy . . . . .	117
5.2.2. System statystyczny . . . . .	118
5.2.3. System sitowy . . . . .	120
5.2.4. System neuronowy . . . . .	122
5.2.5. System hybrydowy . . . . .	125
5.3. Dekodowanie relacji pośrednich i pomocniczych . . . . .	125
<b>6. Ewaluacja szczegółowa</b>	<b>127</b>
6.1. Wykrywanie wzmianek . . . . .	127
6.2. Wykrywanie koreferencji . . . . .	128
6.2.1. Wzmianki idealne . . . . .	128
6.2.2. Wzmianki systemowe . . . . .	129
6.3. Wykrywanie wybranych zależności pośrednich i pomocniczych . . .	130
6.4. Analiza błędów . . . . .	130
6.4.1. Błędy wykrywania wzmianek . . . . .	133
6.4.2. Błędy wykrywania koreferencji . . . . .	135
6.4.3. Analiza relacji pośrednich . . . . .	136
<b>7. Perspektywy badań</b>	<b>139</b>
7.1. W stronę koreferencji uniwersalnej . . . . .	139
7.2. Model Penn Discourse Treebank . . . . .	142
7.3. Anotacja metatekstowa . . . . .	146
<b>Podsumowanie</b>	<b>149</b>
<b>English summary</b>	<b>153</b>
<b>Bibliografia</b>	<b>161</b>
<b>Skorowidz</b>	<b>187</b>
<b>Skorowidz terminów angielskich</b>	<b>189</b>
<b>Wykaz powstałych narzędzi i zasobów</b>	<b>191</b>



---

# Table of contents

<b>Preface</b>	<b>13</b>
<b>Funding information</b>	<b>15</b>
<b>Acknowledgements</b>	<b>17</b>
<b>1. The point of departure</b>	<b>19</b>
1.1. Reference, coreference, anaphora, association . . . . .	19
1.2. Motivation . . . . .	21
1.3. Research objectives . . . . .	22
1.4. Scope of work . . . . .	23
1.5. Methodology . . . . .	24
<b>2. From theoretical perspective to decoding of referential relations</b>	<b>27</b>
2.1. The concept and scope of reference . . . . .	27
2.2. Classifications of mention types and referential relations . . . . .	29
2.2.1. Klemensiewicz . . . . .	30
2.2.2. Topolińska . . . . .	31
2.2.3. Paduczewa . . . . .	32
2.2.4. Clark and other foreign classifications . . . . .	33
2.3. Features of referential relations . . . . .	35
2.4. Corpus projects and automated resolution . . . . .	38
2.5. Computer-based implementations of reference . . . . .	41
2.6. Evaluation methods . . . . .	48
2.6.1. MUC metric . . . . .	50
2.6.2. B <sup>3</sup> metric . . . . .	51
2.6.3. CEAF metric . . . . .	52
2.6.4. BLANC metric . . . . .	53
<b>3. Typology of referential relations</b>	<b>55</b>
3.1. Discourse world and referential properties . . . . .	55
3.2. Mention types and borders . . . . .	56

3.3. Textual relations vs. out-of-text reference . . . . .	58
3.4. Referential typology . . . . .	59
3.4.1. Coreference . . . . .	61
3.4.2. Indirect reference . . . . .	61
3.4.3. Supporting relations . . . . .	63
3.4.4. Excluding relations . . . . .	63
3.4.5. Facets . . . . .	64
<b>4. Corpus of referential relations</b> . . . . .	<b>67</b>
4.1. Text selection . . . . .	67
4.2. Annotation strategy . . . . .	69
4.2.1. Number and profile of annotators . . . . .	69
4.2.2. Serial vs. parallel annotation . . . . .	71
4.2.3. Pre-annotation . . . . .	72
4.2.4. Automated adjudication . . . . .	73
4.3. Annotation phases . . . . .	75
4.3.1. Preparatory phase . . . . .	75
4.3.2. Annotation of nominal coreference . . . . .	76
4.3.3. Annotation of referential relations . . . . .	78
4.4. Annotation tools . . . . .	79
4.5. Annotator agreement . . . . .	83
4.5.1. Mentions . . . . .	83
4.5.2. Coreference clusters . . . . .	84
4.5.3. Other relations . . . . .	85
4.6. Error correction . . . . .	87
4.7. Corpus availability . . . . .	88
4.7.1. SemEval/CoNLL format . . . . .	89
4.7.2. MMAX format . . . . .	91
4.7.3. TEI format . . . . .	94
4.7.4. BRAT format and online corpus version . . . . .	96
4.7.5. Corpus search engine . . . . .	101
4.8. Corpus statistics . . . . .	101
4.8.1. Textual properties . . . . .	101
4.8.2. Mention statistics . . . . .	103
4.8.3. Coreference clusters . . . . .	107

---

<b>5. Implementation</b>	<b>113</b>
5.1. Mention detection . . . . .	113
5.1.1. Rule-based mention detection . . . . .	114
5.1.2. Statistical mention detection . . . . .	115
5.2. Coreference resolution . . . . .	117
5.2.1. Rule-based coreference resolution . . . . .	117
5.2.2. Statistical coreference resolution . . . . .	118
5.2.3. Sieve-based coreference resolution . . . . .	120
5.2.4. Deep network-based coreference resolution . . . . .	122
5.2.5. Hybrid system . . . . .	125
5.3. Decoding associative and auxiliary relations . . . . .	125
<b>6. Evaluation</b>	<b>127</b>
6.1. Mention detection . . . . .	127
6.2. Coreference resolution . . . . .	128
6.2.1. Gold mentions . . . . .	128
6.2.2. System mentions . . . . .	129
6.3. Detection of selected indirect relations . . . . .	130
6.4. Error analysis . . . . .	130
6.4.1. Mention detection errors . . . . .	133
6.4.2. Coreference resolution errors . . . . .	135
6.4.3. Analysis of bridging relations . . . . .	136
<b>7. Research perspectives</b>	<b>139</b>
7.1. Towards Universal Coreference . . . . .	139
7.2. Penn Discourse Treebank model . . . . .	142
7.3. Discourse-based annotation . . . . .	146
<b>Conclusions</b>	<b>149</b>
<b>English summary</b>	<b>153</b>
<b>Bibliography</b>	<b>161</b>
<b>Glossary</b>	<b>187</b>
<b>Glossary of English terms</b>	<b>189</b>
<b>Implemented tools and resources</b>	<b>191</b>