# Vom Bild zum Text und wieder zurück

**Donig, Simon**
simon.donig@uni-passau.de
Universität Passau, Deutschland

**Christoforaki, Maria**
maria.christoforaki@uni-passau.de
Universität Passau, Deutschland

**Bermeitinger, Bernhard**
bernhard.bermeitinger@uni-passau.de
Universität Passau, Deutschland

**Handschuh, Siegfried**
siegfried.handschuh@unisg.ch
Universität St. Gallen, Schweiz

In den letzten Jahren hat die Anwendung von Verfahren der Computer Vision im Bereich der digitalen Kunstgeschichte und Objektforschung erheblich an Bedeutung gewonnen (Donig, Handschuh, Hastik, Kohle, Ommer, Radisch, Rehbein 2018). Dabei stellt das Schließen der semantischen Lücke eine zentrale Herausforderung für (teil-)automatisierte algorithmische Verfahren dar. Hier schlagen wir einen multimodalen Zugang vor, in dem wir eine fruchtbringende Lösung des Problems sehen und den wir im Kontext des Neoclassica-Projekts entwickeln.

Neoclassica ist ein Rahmenwerk zur Erforschung der ästhetischen Kultur des Klassizismus (ca. 1760-1860), das Methoden und Instrumente zur Erforschung von Architektur und Raumkunst bereitstellt (Donig, Christoforaki, Bermeitinger, Handschuh 2017). Dazu bedient es sich eines Ansatzes des Wissensrepräsentation in der Form einer eigenen Ontologie (Donig, Christoforaki, Handschuh 2016) sowie datengetriebener Forschungsinstrumente aus dem Bereich der künstlichen Intelligenz, hier insbesondere der Klassifizierung von Bildern und der semantischen Segmentierung von Bildinhalten mit Verfahren des Deep Learning (Donig, Christoforaki, Bermeitinger, Handschuh 2018).

Algorithmische Werkzeuge bedürfen qualitativer Metriken, um ihre Verlässlichkeit und Reproduzierbarkeit abzubilden. Was aber, wenn die Klassifizierung nicht auf einer Serie flacher Label beruht, sondern wenn die Grundlage für die Klassifizierung komplexe Konzepte sind, die durch semantische Hierarchien verbunden werden wie im Fall einer Annotation von Bilddaten mit einer Ontologie?

Bei unserer Arbeit an Neoclassica sind wir diesem Problem wiederholt begegnet. Wenn wir zum Beispiel einen Armlehnstuhl (Abb.1) in einem Bildwerk annotiert haben, dann weist das Konzept ohne Zweifel Gemeinsamkeiten mit dem eines Stuhls auf. In der Ontologie wird dieser Umstand dadurch ausgedrückt, dass *Stuhl* eine übergeordnete Klasse zur Klasse *Armlehnstuhl* ist (Abb.2).

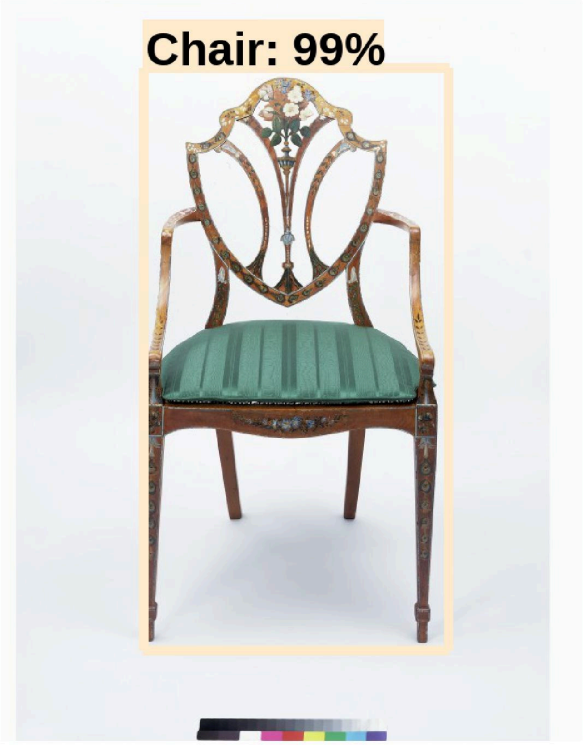


*Abb.1 Die Bounding-box um das Möbel spiegelt die Konfidenzrate des Algorithmus wieder.*

Während für einen menschlichen Beobachter ein Armlehnstuhl eine spezielle Unterkategorie von Stühlen darstellt, ist für die von uns genutzten Algorithmen diese Zuordnung dagegen falsch - bezogen auf die von uns ursprünglich vorgenommenen Annotationen-, da sie die elementare semantische Beziehung zwischen den Konzepten nicht berücksichtigen.

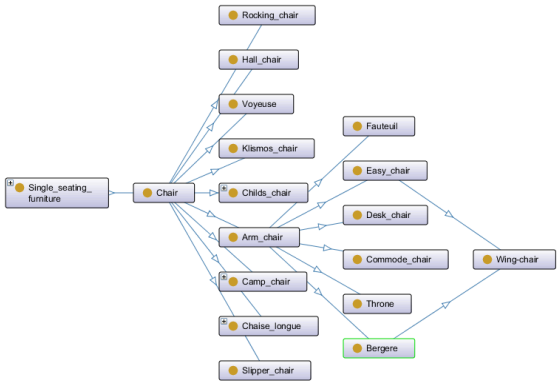

*Abb.2 Neoclassica-Ontologie: Die Chair -Subhierarchie*

Ein vergleichbares Problem stellt der Umgang mit Geschwisterklassen einer formalen Wissensrepräsentation dar. Armlehnstühle mit offenen (*Fauteuil*) und geschlossenen (*Bergère*) Armlehnen sind sich in vielen Aspekten ähnlich. Dennoch ist dem Klassifizierungsprozess dieses Verwandtschaftsverhältnis zunächst nicht zu eigen (Abb.3).


*Abb.3 Geschwisterklasse mit hoher Konfidenzrate klassifiziert*

# Vorgehensweise

Wir stellen hier einen multimodalen Zugang vor, der einen Ansatz aus dem NLP, einem Bereich, wo solche Beziehungen schon lange eingehend studiert worden sind (Indurkhya, Damerau 2010: 120), (Miller 1995), mit einem Ansatz der Computer Vision verbindet - dem Deep Learning visueller Merkmale.

Dieser Zugang beruht auf der *distributional hypothesis*, die postuliert, dass eine Korrelation zwischen der Verteilung von Wörtern und ihrer semantischen Eigenschaften in einem Textkorpus besteht (Rubenstein & Goodenough 1965), was erlaubt, mit Hilfe ersterer die zweiten abzuschätzen (Sahlgren, 2008). Dies schließt somit auch Generalisierungen und Spezialisierungen o.ä. zwischen verschiedenen Klassen ein.

Die ausführlichste systematische Anwendung der Verteilungshypothese findet sich in Distributional Semantic Models (DSMs), die einen multidimensionalen Vektorraum bilden, in dem Wörter als Vektoren abgebildet werden (Lenci, 2018). Diese Vektoren bilden die Kookkurrenz eines Worts mit anderen Wörtern in einem Textkorpus ab, nähern sich so einem Kontext bzw. der Bedeutung dieses Wortes an. Diesen Prozess, in dem ein Wort auf einen 3Vektor abgebildet wird, bezeichnen wir als *word embeddings* (Mikolov, Chen, Corrado, Dean 2013), (Collobert, Weston 2008). Der Grad semantischer Nähe von zwei Wörtern kann 3durch die Anwendung mathematischer Formeln auf diese Vektoren repräsentiert werden (Budanitsky, Hirst 2006).

Für den hier vorgeschlagenen Beitrag haben wir ein DSM beruhend auf einem domänenspezifischen Textkorpus erzeugt.[1] Dazu benutzen wir das an unserem Lehrstuhl entwickelte Indra-Framework, das die Erzeugung, Verwendung und Evaluierung von Word Embedding-Modellen unterstützt (Sales, Souza, Barzegar, Davis, Freitas, Handschuh 2018).

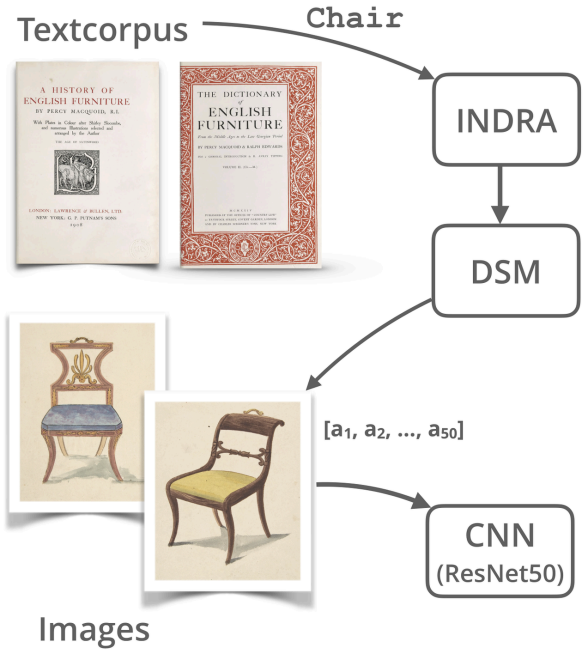


*Abb.4 Trainingsprozess*

Das DSM erlaubt es uns, Word Embeddings für die Klassen in unserem Neoclassica-Open-Korpus (Donig, Christoforaki, Bermeitinger, Handschuh 2018: 131;133) zu erstellen. Für den ersten Schritt dieses Experiments beschränken wir uns auf Bildwerke von einzelnen Objekten.

Anschließend trainieren wir ein Neuronales Netz (ResNet50 (He, Zhang, Ren, Sun 2015)) zur Bildklassifizierung statt mit herkömmlichen, flachen Labels mit den aus dem DSM hervorgehenden Vektoren. In

Abb.4 illustrieren wir den Prozess, bei dem das Wort *Chair* mit dem Vektor [a1, a2, … a50] korrespondiert, der dann genutzt wird, um Bilder von Stühlen zu annotieren.

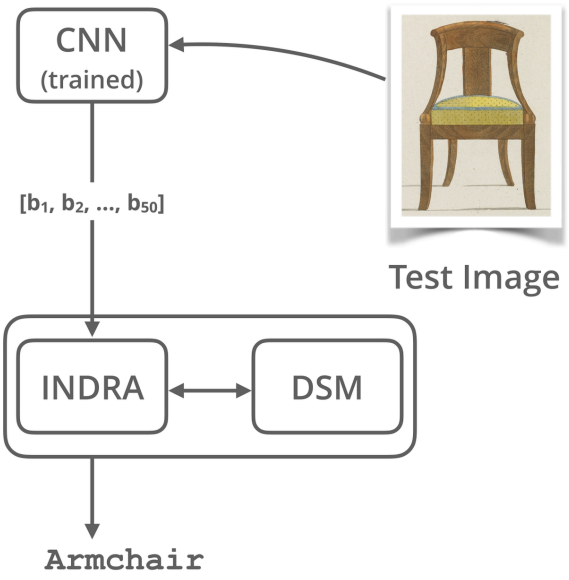


*Abb.5 Testphase*

Die Testphase wird in Abb.5 illustriert, dabei wird ein Testbild in das CNN eingespeist, dass es mit einem Vektor assoziiert.

Indra ermöglicht es uns nun, die nächsten Nachbarn dieses Vektors in Wörtern zu finden und diesen ein Text-Label zuzuordnen. Diese Relationen zwischen den Wörtern stellen zugleich eine semantische Beziehung der Bildinhalte her.

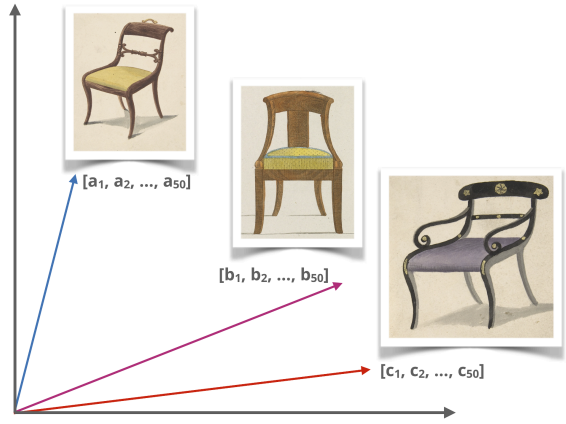


*Abb.6 Idealtypische Repräsentation des reduzierten Vektorraums*

Illustriert wird diese Beziehung in Abb.6, die eine idealtypische Visualisierung des 50-Dimensionalen Vektorraums, reduziert auf zwei Dimensionen, zeigt. Vektor [$a_1$, $a_2$, … $a_{50}$] entspricht dem Begriff *Stuhl*, wohingegen Vektor [$c_1$, $c_2$, … $c_{50}$] dem Begriff *Armlehnstuhl* entspricht. Das Testbeispiel eines Gondelstuhls entspricht Vektor [$b_1$, $b_2$, … $b_{50}$] , der größere semantische Nähe zu [$c_1$, $c_2$, … $c_{50}$] als zu [$a_1$, $a_2$, … $a_{50}$] aufweist, da der Rücken des Gondelstuhls einen armähnlichen Rahmen besitzt, der sich sanft nach vorne neigt und in die Armstützen übergeht.

Wir hoffen, dass diese semantische Beziehung zwischen mehreren Bildern natürlichsprachige Beziehungen zwischen den abgebildeten Artefakten besser reflektiert, als dies herkömmliche “simple” Klassifizierungsprozesse können.

# Stand der Umsetzung & Teilergebnisse

## Bildanalyse

Bislang haben wir Verfahren aus dem Bereich des Deep Learning eingesetzt, um Abbildungen einzelner Möbel (Bermeitinger, Donig, Christoforaki, Freitas, Handschuh 2017) sowie mit Möbelgruppen in Interieuransichten (Donig, Christoforaki, Bermeitinger, Handschuh 2018) zu klassifizieren. Wir konnten dabei zeigen, dass algorithmische Instrumente hervorragend in der Lage sind, Einzelobjekte zu identifizieren (0,94 aMP) - und dies, relativ unabhängig vom Vorliegen in einer bestimmten Technik und Materialität (Fotografie, Gemälde, Zeichnung, Druckgrafik). Für Darstellungen von Mobiliar in Interieurs können wir in unseren Experimenten immer noch gute Ergebnisse vorweisen (aMP 0.53; recall 0.51). Wie eine qualitative Analyse dieser Ergebnisse gezeigt hat, ist die Differenz zum vorausgegangenen Experiment nicht alleine auf die gestiegene Komplexität (z.B. hohe Zahl der Klassen, Überlappung von Objekten im Raum, generell Noise), sondern auch auf zahlreiche nominelle Fehlklassifizierungen zurückzuführen, die aus dem eingangs geschilderten Hierarchie-Problem resultieren.

## Verteilungssemantik

Da es keine allgemeine Methode der Evaluierung eines domänespezifischen DSM gibt (Lenci, 2018), zeigen wir nachstehend, dass das Modell sinnvolle Ergebnisse produziert, wenn man diese mit Weltwissen sowie der Neoclassica-Ontologie vergleicht.

**armchair** :
['armchair', 'upholst',[2] 'sette', 'cane', 'mendlesham']

**settee** :
['sette', 'upholst', 'windsor', 'stool', 'armchair']

Da Begriffe mit sich selbst am nächsten verwandt sind, erscheinen sie an erster Stelle in der Begriffskette, was als ein Zeichen dafür gewertet werden kann, dass das DSM korrekt funktioniert. Beide Möbel gehören zu einer Klasse von gepolsterten Sitzmöbeln (‘*upholst*’) ; in einigen Fällen lagen Polsterungen auch lose auf einem Geflecht auf (‘*cane*’) . Weiter wird deutlich, dass es eine reziproke Beziehung zwischen beiden Begriffen gibt, denn sie referenzieren sich wechselseitig. Das Sofa weist in diesem Korpus außerdem eine enge Nachbarschaft zu einem weiteren Sitzmöbel, dem Hocker (‘stool’) auf. Insgesamt zeigen die Beispiele also bemerkenswerte semantische Nähe und Geschlossenheit.

Ein abschließendes Beispiel mag der Begriff des mehrarmigen Leuchters sein:

**candelabra** :
['candelabra', 'consol', 'torcher', 'girandol', 'candlestick']

Leuchter existieren in klassizistischen Interieurs für gewöhnlich in Paaren. Es macht daher Sinn, dass diese Leuchter auch im DSM als Mehrzahl auftreten ‘*candelabra*’. Für gewöhnlich stehen sie auf einem Möbel oder Kaminsims (daher ‘*consol*’ für einen Konsoltisch). In der Neoclassica-Ontologie hat die Klasse Candelabrum eine Reihe von ihr verwandten Klassen von Leuchtmitteln, die alle unabhängig von ihrem Vorliegen in der Ontologie auch innerhalb des DSM identifiziert worden sind.

Spezifisch sind dies die in der Ontologie auf einer Ebene angesiedelten Klassen *Candlestick*, *Torchere* und die etwas tiefer in der Hierarchie liegende *Girandole* (Abb.7).

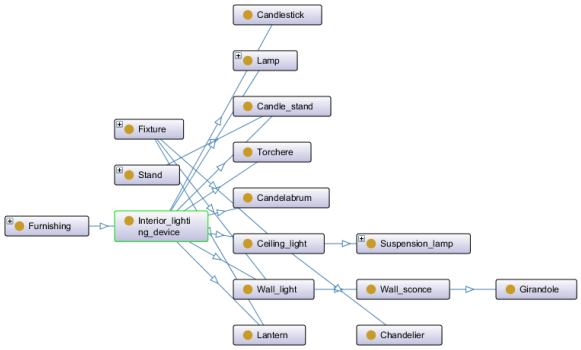


*Abb.7 Neoclassica-Ontologie: Die Interior_lighting_device-Subhierarchie*

An dieser Stelle ist noch einmal zu betonen, dass die Ordnung des DSM rein auf Statistik beruht und aus dem textuellen Korpus abgeleitet ist, während die hier zum Vergleich herangezogene Ordnung der Ontologie eine menschengemachte Wissensrepräsentation ist.

# Conclusum und nächste Schritte

Die vorliegende Kurzzusammenfassung unseres Vortrags schlägt ein Verfahren für eine Einbeziehung semantischer Kontextinformation in den Bildklassifizierungsprozess mit Deep Neural Networks vor. Das Verfahren entsteht im Rahmen des Neoclassica-Projekts und zielt insbesondere darauf, die Erkennung von Mobiliar in historischen Darstellungen von Innenräumen zu verbessern. Der multimodale Zugang wird, so die Hoffnung, dazu beitragen, Schwierigkeiten, denen wir in unserer bisherigen Arbeit begegnet sind - wie der Herausforderung unscharfer Konzepte oder dem Problem der Klassifizierung in semantischen Hierarchien - besser gerecht zu werden. Zukünftige Schritte werden sich auf drei Gebiete erstrecken. Erstens bedarf die Konstruktion eines domäne- und aufgabengerechten DSMs weiterer Verfeinerung. Es gilt zu evaluieren, ob der Umfang des Korpus für das beabsichtigte Ziel bereits ausreichend ist. Qualitätskriterien für eine Evaluierung müssen entwickelt werden, die nicht alleine nach NLP-Maßstäben, sondern auch im Domänezusammenhang sinnvoll sind. Zweitens gilt es eine adäquate Lösung für den Umgang mit zusammengesetzten Ausdrücken zu finden. (Zur Herausforderung semantischer Kompositionalität im Kontext des DSM vgl. Baroni et al., 2014).

Die vorläufigen Ergebnisse der beiden ersten Meilensteine in den Bereichen der Bildanalyse und der Transformation des Textkorpus in ein DSM geben uns die Hoffnung, dass die Einführung von Kontext in den Bildklassifizierungsprozess die *fuzzyness* des Domänegegenstands besser akkomodiert und damit letztlich auch zu einer Verbesserung der Trefferquote des Klassifikationsverfahrens beiträgt.

# Fußnoten

1. Das Textkorpus umfasst 32 Quellen, die 1.987.544 Worte und 58.651 unique word forms repräsentieren. Es besteht aus englischsprachigen Fachpublikationen der Jahrhundertwende vom 19. zum 20. Jahrhundert (cf. Abschnitt Quellen im Literaturverzeichnis). Wir haben diese Texte ausgewählt, da sie stärker differenzierte Konzepte zur Beschreibung des Fachgebiets bieten und die Qualität der von uns durchgeführten optischen Zeichenerkennung (OCR) für diesen Zeitabschnitt deutlich höher war als für zeitgenössische Texte. Anders als moderne Fachtexte sind diese Publikationen zudem unter einer freien, permissiven Lizenz verfügbar.
2. Die Begriffe sind innerhalb des DSMs auf ihren Wortstamm zurückgeführt.

# Bibliographie

**Baroni, Marco / Bernardi, Raffaela / Zamparelli, Roberto (2014)**: *"Frege in Space: A Program of Compositional Distributional Semantics"*, LiLT (Linguistic Issues in Language Technology), 9: 241–346.

**Bermeitinger, Bernhard / Donig, Simon / Christoforaki, Maria / Freitas, André / Handschuh, Siegfried (2017)**: *"Object Classification in Images of Neoclassical Artifacts Using Deep Learning."* Montreal, Canada. https://dh2017.adho.org/abstracts/590/590.pdf [Letzter Zugriff 25.09. 2018]

**Bontempi, Gianluca (2017)**: *Handbook-Statistical Foundations of Machine Learning*. Bruxelles: Machine Learning Group Computer Science Department ULB Belgique.

**Budanitsky, Alexander / Hirst, Graeme (2006)**: *"Evaluating Wordnet-Based Measures of Lexical Semantic Relatedness"*, Computational Linguistics 32 (1): 13–47.

**Collobert, Ronan / Weston, Jason (2008)**: *"A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning"*, in: Proceedings of the 25th International Conference on Machine Learning, 160–167.

**Donig, Simon / Christoforaki, Maria / Handschuh, Siegfried (2016)**: *"Neoclassica - A Multilingual Domain Ontology. Representing Material Culture from the Era of Classicism in the Semantic Web"*, in: **Bozic, Bojan/Mendel-Gleason, Gavin/Debruyne, Christophe / O'Sullivan, Declan (eds.)**: *Computational History and Data-Driven Humanities*. CHDDH 2016 (=IFIP Advances in Information and Communication Technology, vol 482), Cham: Springer: 41–53, DOI 10.1007/978-3-319-46224-0_5.

**Donig, Simon / Christoforaki, Maria / Bermeitinger, Bernhard/ Handschuh, Siegfried (2017)**: *"Neoclassica – an Open Framework for Research in Neoclassicism."* Montreal, Canada. https://dh2017.adho.org/abstracts/384/384.pdf [Letzter Zugriff 25. 09. 2018]

**Donig, Simon / Christoforaki, Maria / Bermeitinger, Bernhard / Handschuh, Siegfried (2018)**: *"Bildanalyse durch Distant Viewing - zur Identifizierung von klassizistischem Mobiliar in Interieurdarstellungen"*, in: **Vogeler, Georg (ed.)**: *DHd 2018 - Kritik der digitalen Vernunft*. Köln: 130–137.

**Donig, Simon / Handschuh, Siegfried / Hastik, Canan / Kohle, Hubertus / Ommer, Björn / Rehbein, Malte (2018)**: *"Der ferne Blick. Bildkorpora und Computer Vision in den Geistes- und Kulturwissenschaften - Stand - Visionen - Implikationen"*, in: **Vogeler, Georg (ed.)**: *DHd 2018 - Kritik der digitalen Vernunft*. Köln: 86–89.

**He, Kaiming / Zhang, Xiangyu / Ren, Shaoqing / Sun, Jian (2016)**: *"Deep residual learning for image recognition"*, in: Proceedings of the IEEE conference on computer vision and pattern recognition: 770-778.

**Indurkhya, Nitin / Damerau, Fred J. (2010)**: *Handbook of Natural Language Processing. Second Edition. Vol. 2. Machine Learning & Pattern Recognition Series*. Boca Raton, FL: Chapman & Hall/CRC Taylor & Francis Group.

**Lenci, Alessandro (2018)**: *"Distributional models of word meaning"*, in: Annual review of Linguistics, 4 (1) :151-171.

**Miller, George A. (1995)**: *"WordNet: A Lexical Database for English"*, in: Communications of the ACM 38 (11): 39–41.

**Sales, Juliano Efson / Souza, Leonardo / Barzegar, Siamak / Davis, Brian / Freitas, André / Handschuh, Siegfried (2018)**: *"Indra: A Word Embedding and Semantic Relatedness Server."* In LREC. Miyazaki, Japan, 2018.

**Mikolov, Tomas / Chen, Kai / Corrado, Greg / Dean, Jeffrey (2013)**: *"Efficient Estimation of Word Representations in Vector Space."* ArXiv:1301.3781 [Cs]. http://arxiv.org/abs/1301.3781. [Letzter Zugriff 25. 09. 2018]

**Rubenstein, Herbert / Goodenough, John B. (1965)**: *"Contextual Correlates of Synonymy."* Communications of the ACM 8 (10): 627-6337. https://doi.org/10.1145/365628.365657. [Letzter Zugriff 25. 09. 2018]

**Sahlgren, Magnus (2008)**: *"The Distributional Hypothesis."* Italian Journal of Linguistics 20, (1): 33–53.