



**SEVENTH FRAMEWORK PROGRAMME
THEME 3
Information and Communication Technologies**

PANACEA Project

Grant Agreement no.: 248064

Platform for Automatic, Normalized Annotation and
Cost-Effective Acquisition
of Language Resources for Human Language Technologies

D5.7

Sample of Transfer Entries produced

Dissemination Level: Public
Delivery Date: June 15th 2012
Status – Version: Final v1.0
Author(s) and Affiliation: Gregor Thurmair, Vera Aleksić (Linguattec)



UNIVERSITY OF
CAMBRIDGE

linguatec
language technologies



Relevant Documents

Panacea Deliverable D5.4: English-French and English-Greek bilingual dictionaries for the Environment and Labour Legislation domains

Panacea Deliverable D5.5: English-French and English-Greek bilingual dictionaries for the Environment and Labour Legislation domains

Panacea Deliverable D5.6: Transfer Selection Support

Panacea Deliverable D7.4: Evaluation Report (Third cycle)

Table of Contents

1. Introduction	4
2. The Conceptual Lexicon	4
2.1 Background	4
2.2 Format	4
3. The Probability Lexicon.....	5
3.1 Background	5
3.2 Format	5
4. Conceptual Lexicon: Sample entries.....	6
5. Probability Lexicon: Sample Entries.....	11

1. Introduction

The present deliverable presents sample entries from tools created in WP 5, task 3: Transfer Rule production.

The report D5-6 documents the approach taken; in short it consists of the following steps:

1. Starting point is an existing bilingual dictionary; it contains, among others, entries with multiple transfers. The challenge is to provide contexts which determine the transfer selection in a given context.
2. For each transfer entry of a given package, a special corpus is built in which the target lemma is the translation of a source lemma. This is basically done by creating subcorpora with matching sentence-IDs and by word-aligning the translation candidates.
3. In each subcorpus, clusters are created on the source language side containing the concepts which co-occur with significant frequency with the translation candidate (source lemma). These clusters are collected and form the Conceptual Lexicon; sample entries of it are shown below.
4. As many entries will not have sufficient context to produce conceptual clusters, a fallback strategy is used to select transfers, consisting in probabilities of different kinds; the translation with best probability score is selected. These probabilities are collected in a second lexicon, called Probability Lexicon; sample entries of it are shown below as well.

Evaluation shows that using these two additional resources can improve the transfer selection process significantly. As it is an independent resource it can be used by any MT system, be it shallow or deep.

2. The Conceptual Lexicon

2.1 Background

The conceptual lexicon is created by analysing the subcorpus attached to each entry for co-occurrences: All lemmata of all sentences of the subcorpus are compared, and the ones with the best co-occurrence score are taken. Preliminary experiments to restrict the clusters produced to a certain size, or to use a threshold, showed that the data sparsity requires to basically leave all candidates in the clusters.

Also, lemmata co-occurring with several transfers of a package were not eliminated but left in the clusters, as they could still help to disambiguate from other translation candidates, and would leave many transfers with very few contexts.

In addition, experiments to include the distance of the co-occurring word to the lemma weight as an additional precision measure have been postponed; their influence would only be marginal¹.

The output of the component which builds the conceptual lexicon is the lexicon itself. The lexicon has about 22.000 entries.

2.2 Format

The lexicon itself gives, for each source language package defined by <lemma, part-of-speech>, a list of translations, consisting of: the translation, its part-of-speech, and an optional cluster of variable size, consisting of pairs of <sl-lemma, weight>, the weight giving the strength of the co-occurrence:

¹ an exception may be certain prepositions indicating strict subcategorisation; to be investigated.

```

ConceptualLexicon ::= entry+
entry              ::= transfer context?
transfer           ::= sourcelemma sourcepos targetlemma targetpos
context            ::= concept (, concept)*
concept            ::= sourcelemma ']' weight

```

File is UTF8, fields are separated by <tab>.

3. The Probability Lexicon

3.1 Background

In case the conceptual transfer does not lead to a result (and this case is rather likely given the amount of transfers without any context because there were no sentences), a fallback strategy is created, which consists in computing a translation probability score.

Previous experiments in the creation of the LinguaDict lexicon have shown that simply using the corpus frequency of a translation is not the best option: We want to know how often the target lemma occurs *as translation of a given source lemma*. Otherwise target lemmata which are very often used as translations (like ‘be’ or ‘have’) disturb the transfer selection.

Also, a relevant factor is *for how many words* a given target lemma is a translation: If a target lemma has high frequency as translation of only one source word, this is much more important than if the frequency results from the fact that it is the translation of e.g. five source words.

Therefore a different approach was taken:

The translation probability consists of three scores, differing in the reference used to compute them. These scores are:

- **Package probability:** probability of a given translation related to the other translations of *this package*. Number of sentences for the given translation DIV total number of sentences in this package. (0 for all transfers without a sentence in the subcorpus);
- **Target probability:** probability of a given translation related to *other source terms* (i.e. for how many SL lemmas is this a possible transfer?) (0 for all terms which are in no package)
- **Corpus probability:** probability that this translation is used at all in the *target language*. Number of occurrences of TL lemma DIV number of lemmata in the total TL corpus.

Querying of the probability lexicon is done sequentially, i.e. if a score is zero then the next ‘weaker’ score is taken. Finally, a corpus probability is nearly always available.

The probabilistic lexicon has also 22.000 entries.

3.2 Format

The lexicon gives the three probabilities for each transfer entry.

```

ProbabililtyLexicon ::= entry+
entry                ::= transfer probabilities
transfer             ::= sourcelemma sourcepos targetlemma targetpos
probabilities        ::= packageprob targetprob corpusprob

```

format of the probability lexicon is:

source lemma \t source-pos \t tl-lemma \t tl-pos \t package prob. \t target-prob \t corpus-prob

File is UTF8, fields are separated by <tab>.

4. Conceptual Lexicon: Sample entries

sl-lemma	pos	tl-lemma	pos	conceptual context
gemein	Ad	common	Ad	europäisch 0.28, geben 0.20, union 0.22, merkmal 0.21, tätigkeitsbereich 0.17, zuständigkeits 0.17, abgegrenzt 0.17, selbständigkeit 0.17, haushaltsmäßig 0.17, verwaltungs 0.17, rechtspersönlichkeit 0.17, rechtsgrundlage 0.17, errichtung 0.17, dezentralisiert 0.17, einrichtung 0.17, „ 0.17, gemeinschaftsagentur 0.17, “ 0.17, bezeichnen 0.17, folgen 0.17
gemein	Ad	mean	Ad	sagen 0.33, denken 0.16, machen 0.16, falsch 0.16, unverlangt 0.16, akt 0.16, freundlichkeit 0.16, tun 0.16, echt 0.16, schwer 0.16, welt 0.16, lieben 0.16, weilen 0.16, beschützen 0.16, felix 0.16, erfahren 0.16, hoffen 0.16, wissen 0.16, klingen 0.16
groß	Ad	great	Ad	risiko 0.05, union 0.05, herr 0.05, land 0.06, groß 0.07, unterschied 0.06, europäisch 0.09, gefahr 0.09
groß	Ad	wide	Ad	unterschied 0.17, kluft 0.17, % 0.07, union 0.07, defizit 0.07, wachstum 0.07, öffnen 0.07, viel 0.07, roma-frau 0.07, mädchen 0.07, ganz 0.07, produkt 0.07, offensichtlich 0.07, hoch 0.07, reichen 0.07, tat 0.07, spanne 0.07, region 0.07, problem 0.07, politisch 0.07
groß	Ad	heavy	Ad	hinweisen 0.5, irland 0.5, derzeitig 0.5, richtwert 0.5, einhalten 0.5, sicher 0.5, weit 0.5, verbessern 0.5, finanziell 0.5, belastung 0.5, vornehmen 0.5, eigentlich 0.5, führen 0.5, ende 0.5, kommission 0.5, arbeitspensum 0.5
groß	Ad	high	Ad	gefahr 0.15, schwer 0.05, arbeitnehmer 0.05, parlament 0.05, % 0.05, möglich 0.07, hoch 0.05, prääsident 0.05, mensch 0.06, europäisch 0.08, hoffnung 0.06, neu 0.05, weilen 0.06, entsprechend 0.05, herr 0.06, auswirkung 0.05, land 0.08, interesse 0.07, jahr 0.08
groß	Ad	big	Ad	groß 0.17, klein 0.11, europa 0.11, europäisch 0.11, scheitern 0.07, bank 0.07, lassen 0.09, neu 0.05, geben 0.06, gehen 0.08, problem 0.06, union 0.06
groß	Ad	deep	Ad	unterschied 0.22, meinen 0.22, frau 0.22, sagen 0.22, interesse 0.22, italien 0.11, frankreich 0.11, linke 0.11, rechte 0.11, bush 0.11, gore 0.11, bezug 0.11, finanzpolitik 0.11, finden 0.11, lösung 0.11, gelingen 0.11, wohl 0.11, gegensatz 0.11, ministerrat 0.11, deutlich 0.11
groß	Ad	capital	Ad	schreiben 1.0, fair 1.0, handel 1.0, kapitel 0.5, enthalten 0.5, dokument 0.5, wort 0.5, erweiterung 0.5, weltmarkt 0.5, international 0.5, vorbild 0.5, bedingung 0.5, entsprechend 0.5, “ 0.5, „ 0.5, label 0.5, machen 0.5, aufmerksam 0.5, wichtig 0.5, — 0.5

D5.7: Sample Transfer Entries



groß	Ad	tall	Ad	blond 0.16, haar 0.12, schlank 0.12, sehen 0.12, arme 0.08, reiche 0.08, reichen 0.08, schauen 0.08, hoch 0.08, ähnlichkeit 0.08, hose 0.08, fällen 0.08, mann 0.08, gerät 0.08, 1,60 0.08, m 0.08
groß	Ad	large	Ad	klein 0.17, europäisch 0.10, land 0.10, groß 0.11, union 0.05, mitgliedstaat 0.05
groß	Ad	much	Ad	europäisch 0.12, groß 0.10, union 0.06, kommission 0.09, parlament 0.06, bevölkerung 0.06, weit 0.06, schaden 0.06, land 0.07, jahr 0.06
dichtung	No	seal	No	integriert 0.33, einreichen 0.16, absaugung 0.16, wirksamkeit 0.16, erforderlich 0.16, einzelheit 0.16, unversehrtheit 0.16, einzeln 0.16, drücken 0.16, einfuellstutzen 0.16, dichtung 0.16, fest 0.16, schließstellung 0.16, verschluss 0.16, gehalt 0.16, position 0.16, sicher 0.16, ausliefern 0.16, einbaufertig 0.16, fetten 0.16
dichtung	No	gasket	No	sem 1.0, semicarbazid 1.0, verschlossen 1.0, dicht 1.0, sterilisation 1.0, (2) 1.0, treibmittel 1.0, kunststoffdichtung 1.0, metalldeckel 1.0, schließe 1.0, glasgefäß 2.0, verwenden 2.0, neu 1.0, erkenntnis 1.0, zerfallen 1.0, azodicarbonamid 2.0, hitzeinwirkung 1.0, herstellung 2.0, aufgeschäumt 1.0
dichtung	No	packing	No	
dichtung	No	poetry	No	einführung 0.5, werk 0.5, bringen 0.25, gang 0.25, zivilgesellschaft 0.25, russisch 0.25, spielen 0.25, katalysatorenrolle 0.25, internet 0.25, massenmedien 0.25, buchdruck 0.25, wichtig 0.25, am 0.25, kapitalismus 0.25, finanzinstrument 0.25, belletristik 0.25, osmanisch 0.25, erbe 0.25, klassisch 0.25, islamisch 0.25
druck	No	squeeze	No	preis 0.37, ausüben 0.25, führen 0.25, fördern 0.25, betreiber 0.25, gemeldet 0.25, endverbraucher 0.25, preisbildungsregel 0.25, teilnehmeranschluss 0.25, ausschließen 0.25, stehen 0.12, inbetriebnahme 0.12, kurz 0.12, gefährden 0.12, reaktor 0.12, brennstoff 0.12, verkauf 0.12, gewinn 0.12, russisch 0.12, uno 0.12
druck	No	pressure	No	ausüben 0.35, europäisch 0.12, international 0.06, regierung 0.07, rat 0.05, union 0.06, land 0.08, parlament 0.07, mitgliedstaat 0.06, stark 0.05, kommission 0.09, politisch 0.08, groß 0.06
druck	No	crush	No	
druck	No	impression	No	
druck	No	leverage	No	ausüben 0.71, iran 0.28, prääsident 0.14, uganda 0.14, tansania 0.14, völkergemeinschaft 0.14, allgemein 0.14, üben 0.14, halten 0.14, schach 0.14, terror 0.14, krieg 0.14, gebrauchen 0.14, dringend 0.14, disney 0.14, vorfall 0.14, setzen 0.14, job 0.14, einigen 0.14, überlegung 0.14
druck	No	printing	No	
druck	No	wring	No	

D5.7: Sample Transfer Entries



druck	No	print	No	schlecht 0.33, ermächtigen 0.33, überlassen 0.33, druckerei 0.33, vorbehalt 0.33, heft 0.33, zusammenstellung 0.33, t2m 0.33, vordruck 0.33, mitgliedstaat 0.33, kostspielig 0.16, vertrieb 0.16, bereitstellung 0.16, foto 0.16, material 0.16, opfern 0.16, text 0.16, schande 0.16, baum 0.16, kohlenstoffreservoir 0.16
druck	No	arm-twisting	No	
einbruch	No	setting	No	
einbruch	No	drop	No	% 0.36, aktienmarkt 0.27, jahr 0.27, verbrauch 0.27, stark 0.27, am 0.18, jung 0.18, plötzlich 0.18, cashflow 0.18, uz 0.18, weitere 0.18, preis 0.18, führen 0.18, verkauf 0.18, rückgang 0.18, fallen 0.18, zeitlich 0.18, leicht 0.18, verringerung 0.18, einzig 0.09
einbruch	No	fall	No	dramatisch 0.30, nachfrage 0.23, dollar 0.15, amerikanisch 0.15, führen 0.15, gering 0.15, rückgang 0.15, hoch 0.15, arbeitslosigkeit 0.15, weltweit 0.15, jahr 0.15, weilen 0.15, erlitt 0.15, bse-krise 0.15, niveau 0.15, deutlich 0.15, bestellung 0.15, produktion 0.15, deutsch 0.15, werft 0.15
einbruch	No	burglary	No	diebstahl 0.75, klein 0.5, autodiebstahl 0.5, schwer 0.5, verbrechen 0.5, einstufen 0.25, schwerverbrechen 0.25, \$ 0.25, schadenshöhe 0.25, graffiti-sprühen 0.25, steuer 0.25, am 0.25, trunkenheit 0.25, tausend 0.25, mal 0.25, lebenslang 0.25, erhalt 0.25, geringfügig 0.25, delikt 0.25, drogenvergehen 0.25
einbruch	No	break-in	No	nehmen 0.25, geiselnhaft 0.25, stunde 0.25, mitarbeiter 0.25, meinen 0.25, büro 0.25, jahr 0.25, führen 0.25, tsarukyan 0.25, schläger 0.25, vertuschungsversuch 0.25, anschließend 0.25, demokrat 0.25, hauptquartier 0.25, neigen 0.25, amerikaner 0.25, nixons 0.25, verstoß 0.25, bürgerlich 0.25, freiheit 0.25
einfall	No	idea	No	jahrzehnt 0.28, halten 0.28, kommen 0.28, langen 0.28, retten 0.14, denken 0.14, dauern 0.14, sowjetkommunismus 0.14, verrotten 0.14, inner 0.14, kreis 0.14, wissen 0.14, regime 0.14, auflösen 0.14, wirklich 0.14, will 0.14, schießen 0.14, eigen 0.14, horn 0.14, erinnern 0.14
einfall	No	invasion	No	vertun 0.33, kongo 0.33, welt-bank 0.33, ausländisch 0.33, spender 0.33, bezahlen 0.33, hälfte 0.33, haushalt 0.33, ugandisch 0.33, regierung 0.33, drittel 0.33, geld 0.33, sinnlos 0.33, kriegshandlung 0.33, uganda 0.33, wahren 0.33, bevölkerungsgruppe 0.33, assyrisch 0.33, menschenrecht 0.33, grundlegend 0.33
einfall	No	incidence	No	

einnahme	No	intake	No	wirkung 0.66, führen 0.66, erzeugnis 0.66, unterschiedlich 0.33, substanz 0.33, unklar 0.33, dosis 0.33, anwendungsdauer 0.33, parallel 0.33, laxativ 0.33, angabe 0.33, polyalkohol 0.33, g 0.33, d 0.33, gemäß 0.33, hinweis 0.33, hersteller 0.33, verwendet 0.33, täglich 0.33, gehören 0.33
einnahme	No	income	No	aufwendung 0.4, ertrag 0.4, d 0.2, geschäftsjahr 0.2, jahresabschluss 0.2, beziehen 0.2, berücksichtigen 0.2, rücksicht 0.2, zeitpunkt 0.2, ausgabe 0.2, somalia 0.2, mitte 0.2, süden 0.2, bevölkerungsgruppe 0.2, einkünfte 0.2, vergessen 0.2, piraterie 0.2, primär 0.2, mittel 0.2, obergesellschaft 0.2
einnahme	No	capture	No	tag 0.66, aufkommen 0.33, gottesstaat 0.33, schreckgespenst 0.33, gerichtshof 0.33, islamisch 0.33, union 0.33, mogadischu 0.33, hauptstadt 0.33, somalisch 0.33, jahr 0.33, lassen 0.33, staatsoberhaupt 0.33, rajoelina 0.33, andry 0.33, erklärt 0.33, armee 0.33, prääsidentenpalast 0.33, am 0.33, märz 0.33
einnahme	No	receipt	No	verbuchen 0.85, erzeugnis 0.85, mitgliedstaat 0.57, auslagerungsmonat 0.42, fall 0.42, transfer 0.42, interventionserzeugnis 0.42, abgangsmitgliedstaat 0.42, wert 0.42, null 0.42, bestimmungsmitgliedstaat 0.42, absatz 0.42, bezeichnet 0.42, preis 0.42, betrag 0.28, (2) 0.28, beihilfe 0.28, kurs 0.28, erhöhen 0.14, konformitätsentscheidung 0.14
einnahme	No	taking	No	
einnahme	No	revenue	No	haushaltsplan 0.33, ausgabe 0.25, betrag 0.22, regelung 0.16, artikel 0.16, zweckgebunden 0.16, haushaltsjahr 0.19, rahmenprogramm 0.11, zentrum 0.11, steuer 0.11, hoechstbetrag 0.11, teil 0.11, gemäß 0.13, veröffentlichung 0.11, geltend 0.11, organ 0.11, betreffend 0.11, haushaltsordnung 0.11, eingehen 0.11, wiederverwendung 0.11
streuen	Vb	sprinkle	Vb	gratinieren 1.0, gourmetbrenner 1.0, servieren 1.0, zucker 1.0
streuen	Vb	grit	Vb	
streuen	Vb	spread	Vb	risiko 0.5, krieg 0.25, innovation 0.25, jahr 0.25, % 0.25, nutzen 0.25, breit 0.25, verschieden 0.25, schuld 0.12, übermäßig 0.12, systemisch 0.12, bestehen 0.12, verstärkt 0.12, krise 0.12, resultat 0.12, spekulativ 0.12, blase 0.12, exzessiv 0.12, fremdfinanzierung 0.12, schwerwiegend 0.12
kürzen	Vb	shorten	Vb	kommission 0.57, dauer 0.42, ias 0.28, (1) 0.28, lage 0.28, voraussichtlich 0.28, entwicklung 0.28, olivenölmarkt 0.28, verfahren 0.28, artikel 0.28, verordnung 0.28, nummer 0.28, beschließen 0.28, laufend 0.28, vertrag 0.28, bemühen 0.14, ausdrucksweise 0.14, klar 0.14, verzögerung 0.14, kommen 0.14
kürzen	Vb	pare	Vb	

D5.7: Sample Transfer Entries



kürzen	Vb	cut	Vb	ausgabe 0.21, rat 0.15, europäisch 0.10, % 0.14, mittel 0.14, regierung 0.06, union 0.06, einfach 0.05, euro 0.07, jahr 0.08, parlament 0.07, programm 0.05, gehen 0.06, sagen 0.07, land 0.05, kommission 0.07, mitgliedstaat 0.09
kürzen	Vb	curtail	Vb	ausgabe 0.33, politisch 0.33, beitrag 0.33, modulation 0.33, % 0.33, kommission 0.33, fallen 0.16, staatseinnahme 0.16, kreml 0.16, hauptfetisch 0.16, putin 0.16, umfragewert 0.16, bedenken 0.16, erweisen 0.16, rüsten 0.16, öffentlich 0.16, erhöhen 0.16, auslandshilfe 0.16, versprechen 0.16, erwähnen 0.16
kürzen	Vb	reduce	Vb	kommission 0.25, rat 0.20, mittel 0.19, vorschlag 0.13, parlament 0.12, europäisch 0.14, mitgliedstaat 0.12, ecu 0.07, quote 0.05, haushalt 0.05, ausgabe 0.05, union 0.05, frau 0.05, % 0.06, meinen 0.06, million 0.08, euro 0.07, betrag 0.09, stehen 0.05, zahlung 0.07
spritzen	Vb	spray	Vb	vorteil 0.5, ökologisch 0.5, brauchen 0.5, insektizid 0.5, bewahren 0.5, knollenfäule 0.5, kraut 0.5, heimatland 0.5, großbritannien 0.5, geben 0.5, biokartoffel 0.5, kaufen 0.5, weilen 0.5
spritzen	Vb	splash	Vb	s 1.0, tun 0.5, rumplantschen 0.5, nass 0.5, schutzbrille 0.5, visier 0.5, tragen 0.5, augenschutz 0.5, blut 0.5, körperflüssigkeit 0.5, fliegend 0.5, kontaminiert 0.5, abfall 0.5, gewebe 0.5, gesicht 0.5
spritzen	Vb	shoot	Vb	
spritzen	Vb	inject	Vb	verringern 1.0, nervosität 1.0, reizbarkeit 1.0, gerade 1.0, versuchen 1.0, insulinstift 1.0
spritzen	Vb	squirt	Vb	eltern 1.0, liebesbrief 1.0, zerrissen 1.0, junge 1.0, werfen 1.0, undicht 1.0, puff 1.0, kaputt 1.0, naht 1.0

5. Probability lexicon: Sample Entries

sl-lemma	pos	tl-lemma	pos	package-prob	target-prob	corpus-prob
gemein	Ad	common	Ad	0.9444444444444444	0.004175399663009082	6,43E+11
gemein	Ad	mean	Ad	0.0555555555555555	3,29E+10	5,36E+10
groß	Ad	great	Ad	0.3001644736842105	0.008449430814120742	5,37E+11
groß	Ad	wide	Ad	0.023026315789473683	9,78E+11	9,46E+10
groß	Ad	heavy	Ad	0.001644736842105263	5,55E+11	4,43E+11
groß	Ad	high	Ad	0.078125	0.007405580898368471	4,22E+12
groß	Ad	big	Ad	0.18996710526315788	9,90E+10	1,92E+08
groß	Ad	deep	Ad	0.007401315789473684	0.0012575514733078536	5,14E+10
groß	Ad	capital	Ad	0.001644736842105263	8,22E+09	1,63E+12
groß	Ad	tall	Ad	0.019736842105263157	1,27E+11	1,67E+09
groß	Ad	large	Ad	0.3248355263157895	0.002210989191632762	3,59E+10
groß	Ad	much	Ad	0.05345394736842105	0.0017096124604446636	3,74E+11
dichtung	No	seal	No	0.5454545454545454	6,71E+09	1,16E+11
dichtung	No	gasket	No	0.0909090909090909	2,68E+09	2,21E+09
dichtung	No	packing	No	0.0	0.0	1,05E+09
dichtung	No	poetry	No	0.3636363636363636	1,07E+10	6,15E+08
druck	No	squeeze	No	0.0017909111260353706	2,41E+10	2,31E+10
druck	No	pressure	No	0.9952988582941572	0.0011929283977329531	7,59E+08
druck	No	crush	No	0.0	0.0	1,09E+10
druck	No	impression	No	0.0	0.0	2,76E+11
druck	No	leverage	No	0.0015670472352809493	2,01E+11	5,35E+10
druck	No	printing	No	0.0	0.0	0.0
druck	No	wring	No	0.0	0.0	4,99E+08
druck	No	print	No	0.001343183344526528	2,95E+09	2,40E+11
druck	No	arm-twisting	No	0.0	0.0	0.0
einbruch	No	setting	No	0.0	2,68E+09	1,96E+09
einbruch	No	drop	No	0.34375	7,19E+10	1,53E+11
einbruch	No	fall	No	0.40625	4,63E+12	1,71E+10
einbruch	No	burglary	No	0.125	1,07E+10	2,69E+08
einbruch	No	break-in	No	0.125	1,07E+10	7,68E+07
einfall	No	idea	No	0.7	0.001045086844167758	1,50E+12
einfall	No	invasion	No	0.3	1,61E+10	6,60E+09
einfall	No	incidence	No	0.0	1,45E+11	8,11E+08
einnahme	No	intake	No	0.0555555555555555	2,31E+11	7,34E+09
einnahme	No	income	No	0.09259259259259259	3,62E+11	1,00E+12
einnahme	No	capture	No	0.0555555555555555	4,02E+09	3,26E+10
einnahme	No	receipt	No	0.12962962962962962	3,13E+11	4,71E+11
einnahme	No	taking	No	0.0	0.0	0.0
einnahme	No	revenue	No	0.6666666666666666	1,53E+10	7,36E+10
kürzen	Vb	shorten	Vb	0.025179856115107913	6,56E+10	3,88E+09
kürzen	Vb	pare	Vb	0.0	0.0	3,65E+09

D5.7: Sample Transfer Entries



kürzen	Vb	cut	Vb	0.49640287769784175	5,21E+11	8,29E+10
kürzen	Vb	curtail	Vb	0.02158273381294964	2,26E+11	4,36E+09
kürzen	Vb	reduce	Vb	0.4568345323741007	0.004185458594700569	3,11E+12
spritzen	Vb	spray	Vb	0.3333333333333333	2,15E+10	2,45E+08
spritzen	Vb	splash	Vb	0.3333333333333333	2,15E+10	2,11E+08
spritzen	Vb	shoot	Vb	0.0	3,33E+11	2,63E+10
spritzen	Vb	inject	Vb	0.1666666666666666	2,15E+10	4,55E+09
spritzen	Vb	squirt	Vb	0.1666666666666666	1,07E+10	6,72E+07
streuen	Vb	sprinkle	Vb	0.1111111111111111	2,15E+10	1,54E+09
streuen	Vb	grit	Vb	0.0	0.0	2,59E+09
streuen	Vb	spread	Vb	0.8888888888888888	1,83E+12	1,19E+10