I. Research articles

DOI: 10.17951/et.2016.28.117

Izabela Gatkowska
(Jagiellonian University in Kraków, Poland)

Dom 'house/home' in empirical lexical networks*

The article explains the concept of the empirical lexical network. A comparison is drawn between the meanings of the Polish lexeme *dom* 'house/home' as represented in the Wortschatz corpus-derived network (Univeristy of Leipzig) and the experimental network of the Department of Computational Linguistics of Jagiellonian University in Kraków (DCL JU). In both networks, the most vital meanings are 'family home – place for the family' and 'dwelling place'. Both networks also contain the meaning 'building', whereas the meaning 'institution' is better represented in the corpus-based network. The experimental network additionally contains the meaning 'my shelter'. That network, being characterised by directional internal links, involves subnetworks that explain meanings, e.g. the subnetwork for the meaning 'family home – place for the family' points to the special role of the node *matka* 'mother', which organises the subnetwork that represents that meaning. A comparison of the meanings being identified and explained in specific subnetworks with dictionary definitions suggests that research on the network representation of meaning may be useful in lexicography.

Key words: linguistics, empirical network, lexical semantics, network representation of meaning, house/home

* The article appeared in Polish as "Dom w empirycznych sieciach leksykalnych" in *Etnolingwistyka* 28, pp. 117–135. The present English translation has been financed by the Ministry of Science and Higher Education, project titled "English edition of the journal *Etnolingwistyka. Problemy języka i kultury* in electronic form" (no. 3bH 15 0204 83).

1. Empirical networks

The idea of representing the lexicon of natural language as a lexical network was conceived in the 1970s among psychologists (Kiss et al. 1973) and, simultaneously, in research on artificial intelligence (Schank 1975; Minsky 1975). A network is a structure composed of lexical units and relations between them (Minsky 1975). It is generally assumed that the relations (links) between a lexical item and other lexical items within the network represent the dependencies between the meanings of units thus linked (Clark 1970).

An empirical lexical network can be constructed with the aid of an experiment with a sufficient number of subjects, who provide the first word that comes to mind upon hearing a stimulus word (Kent and Rosanoff 1910) or with the use of an algorithm counting and interpreting the co-occurrence of words in a sufficiently large collection of texts (Biemann et al. 2007). In this way, we obtain an empirically motivated collection of "defined unit–defining unit" pairs. We accept that the meanings of both lexical units are joined though a link measured as the number of co-occurrences of both elements in the set of all answers provided for the defined unit or in the set of all sentences containing it. We will try to show that the strength of the link can help recognise the meanings of the defined units, as well as indicating the hierarchy of meanings in the lexicon – these functions are fulfilled by the strongest links.

The set of links between lexical units (network nodes) constitutes a lexical network. An experimentally constructed network has certain formal properties, absent from the network extracted from corpora of texts: links in an experimental network are directional, always from the stimulus to the response. An analysis of the directionality shows that the network includes subnetworks that explain the meanings identified through link strengths.

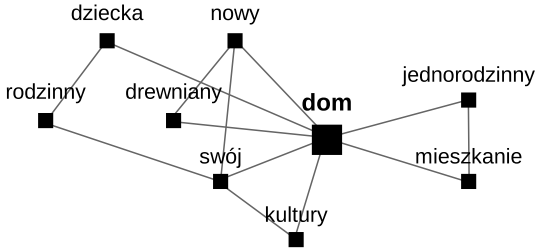
The method of defining meanings in an empirical network will be exemplified with the Polish lexeme *dom* 'house/home'. Then the meanings identified in the network will be compared to those identified in a dictionary compiled according to the rules of the age-long lexicographic tradition.

1.1. Corpus-derived lexical network

We have based our considerations of the network approach to the word *dom* on the Polish corpus and the network-generating algorithms of the Wortschatz project pursued for over a decade now at Leipzig University.[1] Wortschatz includes five Polish-language corpora, three of which are composed of press releases from the years 2007, 2008 and 2011, and two of

[1] http://corpora.informatik.uni-leipzig.de/; access Jan 9, 2016.

Wikipedia entries from the years 2007 and 2010. The lexical network was generated from the corpus of news releases from 2011, containing 6,494,575 sentences and 96,476,260 inflectional forms. Network-generating algorithms function in two phases. First, in operating on sentences, they generate word pairs of the type "defining word–co-occurring word" (e.g. *dom* 'home' – *rodzinny* 'family' (attrib.), i.e. 'family home') and calculate the frequency of occurrence of the pair (i.e. the frequency of co-occurrence of its elements). The pair-generating algorithm is based on the assumption that the co-occurring word must link directly, in a sentence, with the defined word. The lexeme *dom* occurs 8,897 times in the corpus and the list of words co-occurring sufficiently often with *dom* contains 814 items, 407 to the left of it and as many to the right. There are words that occur on both sides, which means that their linkage with the defined word in non-directional. The word *dom* is thus linked with 59 words, which we will call defining words. The most frequent are: *rodzinny* 'family' (attrib.), *mieszkanie* 'apartment, flat', *swój* 'one's own', *nowy* 'new', *drewniany* 'wooden'. The algorithm generating the lexical network of the defined word takes into account, apart from the defining words, also those that appear on both sides of the defining words. As a result, the lexical network constructed around the word *dom* consists of a few hundred word-nodes and a several times greater number of links (dependencies). Because its diagrammatic representation would have been unreadable, we only provide here the network consisting of those words that co-occur with *dom* most frequently (Figure 1).



dziecka 'of a child' (*dom dziecka* 'orphanage)
nowy 'new'
rodzinny 'family' (attrib.)
drewniany 'wooden'
jednorodzinny 'detached; for one family'
swój 'one's own'
kultury 'of culture' (*dom kultury* 'community centre')
mieszkanie 'apartment, flat'

Figure 1. *Dom* in the corpus-derived network

For the purpose of a comparison of a corpus-based network with an experimental network that we propose below, we must recalculate the numerically expressed frequency of occurrence in the corpus into what we call *linkage strength*, expressed as a percentage. This is obtained by dividing

the frequency of co-occurrence of a given defining word by the number of all occurrences of the defined word in the corpus. For example, if the word *rodzinny* 'family' (attrib.) co-occurs with *dom* 1,680 times, and the total number of occurrences of *dom* in the corpus is 8,897, the linkage strength of *dom – rodzinny* equals 18.88% (1,680 ÷ 8,897 ≈ 0.1888, rounded down). We can thus say that the network above consists of the words most strongly linked with *dom*.

1.2. Experimental network at the Department of Computation Linguistics, Jagiellonian University in Kraków (DCL JU)

The experimental network of lexical links arises from the human associative mechanism: the subject provides the first word that comes to mind upon hearing (and understanding) the stimulus word. If we administer this kind of test to a sufficient number of subjects, we receive a list of words linked with a given stimulus word: technically speaking this will be a list of "stimulus–response" pairs. Each pair has an experimentally established strength of linkage between the stimulus and the response; the strength is expressed by the number of subjects that have provided that response. In accordance with the tradition of describing the results of the experiment (Kiss et al. 1973), linkage strength will be expressed as percentage, by dividing the number of the respondents that have provided a given reaction by the total number of respondents. For example, if the number of all responses (reactions) to stimulus A is 95, of which response B was provided 49 times, then the A → B linkage strength equals 52% (49 ÷ 95 rounded up).[3]

If, however, we extend the experiment into a cyclical procedure,[4] i.e. when the responses obtained in the first cycle are used as prompts in the second cycle, we will as a result obtain a different lexical network than that constructed on the basis of corpora. The differences become apparent when we compare Figure 1 with Figure 2, which diagrams the experimental network for the word *dom* (cf. the Appendix for a description of the relevant experiment).

The differences result from the fact that all links in the experimental network are directional and proceed from the stimulus to the response. Thus, the word that acts as a node in a network constructed from stimulus–response pairs may have *outgoing links* (e.g. *dom* → *rodzinny* 'family' (attrib.), where *dom* is the stimulus), as well as *incoming* ones (*drewniany* wooden' → *dom*, where *dom* is the response). In consequence, we can distinguish full

[3] A list of word pairs thus obtained was used in psychiatry as a diagnostic pattern (Kent and Rosanoff 1910).

[4] The mechanism was used for the first time by the authors of *The Edinburgh Associative Thesaurus* (EAT) (Kiss et al. 1973).

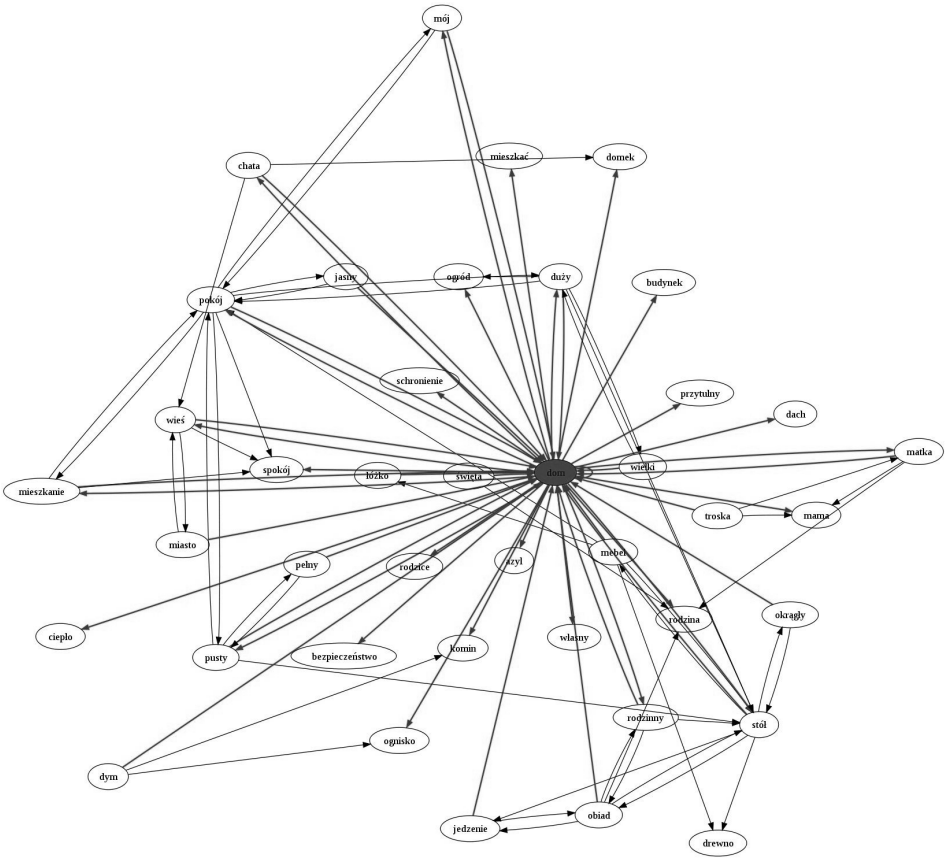


Figure 2. *Dom* in an experimental network. (Explanations of the Polish terms can be found in Table 1.)

nodes, i.e. the words that are connected to others through both outgoing and incoming links (e.g. *dom*) and reduced nodes, i.e. the words that only function in incoming links (e.g. *rodzina* (n.)/*rodzinny* (adj.) 'family'). It is also possible for two words (network nodes) to be connected through a bidirectional, *reciprocal link*, e.g. *dom* → *mieszkanie* 'apartment, flat' or *mieszkanie* → *dom*. These are not frequent cases, e.g. in the experimental network presented in Gatkowska (2014), out of the total of 50,849 links, only 1,181 are reciprocal. We show below that reciprocal links usually participate in subnetworks.

2. *Dom*: the meanings identified through the strongest links

We have shown the corpus/text-based network and the experimental network to have different structures. However, the fact that both networks are

built from word pairs to which linkage strength has been assigned facilitates their comparison. In Table 1, we juxtapose direct links of *dom* in the corpus-generated network, where the links are non-directional, with those in the experimental network, where outgoing and incoming links are distinguished. The items are listed in decreasing-strength order of links with *dom*. The list of words actually linked with *dom* in each network contains hundreds of items – only the strongest links have been selected for comparison.

Table 1. *Dom* in the corpus-generated vs. the experimental networks

Corpus		DCL JU experiment			
		incoming links		outgoing links	
rodzinny 'family' (attrib.)	18.88	*rodzinny* 'family' (attrib.)	16.86	*rodzinny* 'family' (attrib.)	17.67
mieszkanie 'apartment, flat'	14.86	*mieszkanie* 'apartment, flat'	15.25	*mieszkanie* 'apartment, flat'	12.25
swój 'one's own'	10.44	*rodzina* 'family' (n.)	12.16	*mój* 'my'	10.55
nowy 'new'	9.28	*spokój* 'peace and quiet'	2.98	*drewniany* 'wooden'	7.49
drewniany 'wooden'	8.74			*pusty* 'empty'	6.11
jednorodzinny 'detached'[i]	8.04	*ciepło* 'warmth'	2.64	*wielki* 'huge'	5.45
dziecka lit. 'of a child'[ii]	7.78	*ogród* 'garden'	2.41	*chata* 'hut'	4.18
mój 'my'	7.74	*mój* 'my'	2.18	*duży* 'big'	3.45
kultury lit. 'of culture'[iii]	6.92	*dach* 'roof'	1.95	*pokój* 'room'	2.58
kw. 'sq.'[iv]	6.87	*mama* 'mom'	1.61	*wysoki* 'tall, high'	2.18
pow. 'living area'	5.84	*bezpieczeństwo* 'security'	1.49	*ulica* 'street'	2.00
domu 'of the house/home'	5.68			*piękny* 'beautiful'	1.64
którym 'which'	5.24	*duży* 'big'	1.49	*obiad* 'lunch; dinner'	1.49
kupił '(he) bought'	5.00	*pokój* 'room'	1.38	*mały* 'small'	1.42
mieszkalny lit. ' for living'[v]	4.89	*pusty* 'empty'	1.38	*spokojny* 'peaceful, quiet'	1.31
		dom 'house/home'	1.15		
własny 'one's own'	4.62	*zły* 'bad'	1.15	*zły* 'bad'	1.27
wart 'worth'	4.48	*budynek* 'building'	1.03	*biały* 'white'	1.24
rodzinę 'family' (n., Sing Acc)	4.45	*chata* 'hut'	0.92	*ładny* 'nice'	1.24
		własny 'one's own'	0.80	*mebel* 'furniture'	1.24
m. 'metres'[vi]	4.35	*matka* 'mother'	0.69	*dym* 'smoke'	1.16
uderzył '(it) struck' (masc.)	4.00	*mieszkać* 'live, dwell'	0.57	*okno* 'window'	1.09
		rodzice 'parents'	0.57	*wieś* 'countryside; village'	0.87
		azyl 'refuge'	0.46		
		domek 'cottage'	0.46	*wuja Toma* 'uncle Tom's'	0.87
		komin 'chimney'	0.46		
				matka 'mother'	0.73

[i] lit. 'for one family'
[ii] in the collocation *dom dziecka* 'orphange'
[iii] in the collocation *dom kultury* 'community centre'
[iv] as in *m. kw.* 'square metres'
[v] in the collocation *dom mieszkalny* 'dwelling house'
[vi] as in *m. kw.* 'square metres'

Corpus		DCL JU experiment			
		incoming links		outgoing links	
dzieci 'children'	3.97	*ognisko* 'hearth and home'	0.46	*podłoga* 'floor' (ground)	0.73
nasz 'our'	3.96	*przytulny* 'cosy'	0.46	*pełny* 'full'	0.69
wybudować 'build'	3.84	*schronienie* 'shelter'	0.46	*praca* 'work'	0.65
samochód 'car'	3.79	*stół* 'table'	0.46	*ziemia* 'land'	0.51
spłonął '(it) burnt down'	3.78	*wieś* 'countryside; village'	0.46	*za wsią* 'outside the village'	0.47
uderzyła '(it) struck' (fem.)	3.73	*ciepły* 'warm'	0.34	*święta* 'holidays'	0.40
wybudował '(he) built'	3.72	*daleko* 'far away'	0.34	*dom* 'house/home'	0.36
kupić 'buy'	3.52	*dziecka* lit. 'of a child'ii	0.34	*polski* 'Polish'	0.36
towarowy lit. 'with goods'[vii]	3.46	*miłość* 'love'	0.34	*żelazny* 'iron' (attrib.)	0.36
tys. 'thousand'	3.45	*otwarty* 'open'	0.34	*kolorowy* 'colourful'	0.33
gdzie 'where'	3.42	*Tom* (name)	0.34	*dywan* 'carpet'	0.29
TIR 'TIR vehicle'	3.42	*wielki* 'huge'	0.34	*jasny* 'bright'	0.29
stoi 'stands'	3.34	*cegła* 'brick'	0.23	*łóżko* 'bed'	0.29
działkę 'plot' (Sing Acc)	3.32	*chałupa* 'dilapidated house'	0.23	*miasto* 'city'	0.29
ogród 'garden'	3.28	*chatka* 'hut'	0.23		
mody lit. 'of fashion'[viii]	3.23	*drewno* 'wood'	0.23		
pod 'under'	3.14	*drzwi* 'door'	0.23		
wjechał '(he) drove into'	3.14	*jednorodzinny* 'detached'i	0.23		
maklerski lit. 'stockbroker's'[ix]	2.99	*miejsce* 'place'	0.23		
spalił '(it/he) burnt down'	2.96	*obiad* 'lunch; dinner'	0.23		
ogrodem 'garden' (Sing Instr)	2.80	*osiedle* 'housing estate'	0.23		
handlowy 'to do with trade'[x]	2.60	*ostoja* 'anchor, linchpin'	0.23		
gospodarstwo 'homestead'	2.45	*podwórko* 'backyard'	0.23		
i 'and'	2.43	*powrót* 'return' (n.)	0.23		
powierzchni 'of living area'	2.42	*radość* 'joy'	0.23		
stał 'stood'	2.37	*stan* 'state, condition'	0.23		
zbudować 'build'	2.29	*szczęście* 'happiness'	0.23		
aukcyjny 'auction' (attrib.)[xi]	2.24	*ściany* 'walls'	0.23		
		twierdza 'castle'	0.23		

[vii] in the collocation *dom towarowy* 'department store'
[viii] in the collocation *dom mody* 'fashion house'
[ix] in the collocation *dom maklerski* 'brokerage (firm)'
[x] in the collocation *dom handlowy* 'department store'
[xi] in the collocation *dom aukcyjny* 'auction house'

Corpus		DCL JU experiment	
		incoming links	outgoing links
zniszczył '(it/he) destroyed'	2.20		
stary 'old'	2.18		
ciężarówka 'lorry'	2.18		
ma 'has'	2.15		
duży 'big'	2.15		
kupili '(they) bought'	2.11		
zakonny 'monastic'[xii]	2.11		
drewniany 'wooden'	2.10		

[xii] in the collocation *dom zakonny* 'monastic house; monastery'

The table shows a striking convergence: the words *rodzinny* 'family' (attrib.) and *mieszkanie* 'apartment, flat' are the most strongly linked with *dom* in both networks.[7] Additionally, both words enter into reciprocal links. One can thus assume that the strongest links concern two meanings of *dom*. The first is 'the family home, the place for the family to live', which is also signalled by the following weaker links: *rodzina* 'family' and *dzieci* 'children' in the corpus-based network, as well as the outgoing *rodzina* 'family', *mama* 'mom', *matka* 'mother', *rodzice* 'parents' and the incoming *matka* 'mother' in the experimental network. The other salient meaning of *dom* is 'dwelling place', which apart from *mieszkanie* 'apartment, flat' is also signalled by the more weakly linked *kupił* '(he) bought', *własny* 'one's own', *kupić* 'buy', *ma* 'has' in the corpus-based network, plus by the outgoing *pokój* 'room', *własny* 'one's own', *miejsce* 'place', *pusty* 'empty' and *incoming* pusty 'empty', *pokój* 'room', *mebel* 'furniture', *podłoga* 'floor' (ground), *dywan* 'carpet', *łóżko* 'bed' in the experimental network. The only verb that appears as linked in the experimental network with the *dom* node is *mieszkać* 'live, dwell', which is consistent with ethnolinguistic observations (Bartmiński 2015). The other meanings of *dom* are not signalled through a single word that would be strongly linked with it. They can be, however, identified through a group of words more weakly linked, such as the pronoun *mój* 'my', present in both networks and reciprocally linked with *dom* in the experimental network. *Mój* can be associated with outgoing links in the experimental network, such as *spokój* 'peace and quiet', *ciepło* warmth', *bezpieczeństwo* 'security', *własny* 'one's own', *azyl* 'refuge', *ostoja* 'anchor, lynchpin", *schronienie* 'shelter', *szczęście* 'happiness', *twierdza* 'castle' and *przyjąć* 'accept', in that the

[7] Similar results have been obtained in associative experiments, cf. Kurcz 1967 (1,000 respondents): *dom*: *rodzinny* 'family' (attr.) (48), *mieszkanie* 'apartment, flat' (45), *mój* 'my' (45); Gawarkiewicz 2008 (500 respondents): *dom*: *rodzina* 'family' (n.) (164), *rodzinny* 'family' (attr.) (11), *mieszkanie* 'apartment, flat' (22).

meaning of *dom* that the words signal can be described as 'my place, my shelter'.

The next meaning of *dom*, i.e. 'house, building', is represented in both networks. In the corpus-based network the meaning is represented above all by the verbs *kupił* '(he) bought', *uderzył/uderzyła* '(it) struck', *wybudować/zbudować* 'build', *spłonął* '(it) burnt down', *wybudował* '(he) built', *kupić* 'buy', *stoi* 'stands', *wjechał* '(he) drove into', *spalił* '(it/he) burnt down', *stał* 'stood', *zniszczył* '(it/he) destroyed', *ma* 'has', *kupili* '(they) bought', as well as *drewniany* 'wooden', *jednorodzinny* 'detached; for one family', *mieszkalny* (*dom mieszkalny* 'dwelling house'), *samochód* 'car', *TIR* 'TIR vehicle', *działkę* 'plot' (Sing Acc), *ogród* 'garden', *pod* 'under', *gospodarstwo* 'homestead', *ciężarówka* 'lorry'. In the experimental network the meaning 'house, building' is represented by outgoing links: *ogród* 'garden', *dach* 'roof', *budynek* 'building', *domek* 'cottage', *komin* 'chimney', *wieś* 'countryside, village', *cegła* 'brick', *chałupa* 'dilapidated house', *chatka* 'hut', *drewno* 'wood', *jednorodzinny* 'detached; for one family', *drzwi* 'door', *okno* 'window', *osiedle* 'housing estate', *podwórko* 'backyard', as well as the incoming ones: *drewniany* 'wooden', *wielki* 'huge', *dym* 'smoke', *chata* 'hut', *wysoki* 'tall, high', *ulica* 'street', *mały* 'small', *wise* 'countryside, village', *za wsią* 'outside the village', *miasto* 'city'. The last meaning that appears in lexical networks is 'institution', which in the corpus-based network is represented by fixed expressions *dom kultury* 'community centre' (lit. 'house/home of culture'), *dom towarowy* 'department store' (lit. 'house of goods'), *dom mody* 'fashion house', *dom maklerski* 'brokerage (firm)', *dom dziecka* 'orphanage' (lit. 'child's home'), *dom zakonny* 'monastic house', and in the experimental network only by the outgoing link *dom dziecka* 'orphanage'.

We have identified these meanings of the word *dom* intuitively. A justification for the intuition can be found in the empirical network, where the salient meanings surface as clearly recognisable subnetworks. Because elements of the subnetworks are not visible in the simplified diagram for *dom*, we will present a diagram for each subnetwork (for each of the meanings), relating it, where justified, to other network substructures.

3. Subnetworks for distinct meanings of *dom* in the experimental network

In the experimental network, specific meanings of *dom* are organised as subnetworks naturally centred around the *dom* node. The skeleton of a subnetwork are usually reciprocal links between nodes. We assume that subnetworks explain specific meanings of the defined word, while the strengths

of the links that constitute the subnetwork show the importance of the individual senses.

3.1. Meaning: 'family home – place for the family'

The subnetwork that represents the sense 'family home – place for the family' is organised through two very strong reciprocal links, i.e. *dom* ↔ *rodzinny* 'family' (attrib.) (*dom* → *rodzinny* 16.86; *rodzinny* → *dom* 17.67) and several weaker ones, i.e. *dom* ↔ *matka* 'mother' (*matka* → *dom* 0.73; *dom* → *matka* 0.69) and *dom* ↔ *stół* 'table' (*dom* → *stół* 0.46; *stół* → *dom* 0.18), *dom* ↔ *obiad* 'lunch; dinner' (*dom* → *obiad* 0.23; *obiad* → *dom* 1.49). The *rodzinny* ('family', attrib.) node will not be discussed in detail now because it functions as a distinct subnetwork and only connects with the sense 'family home – place for the family' through its links with the nodes *matka* 'mother', *stół* 'table' and *rodzina* 'family' (n.), which define the sense in question.

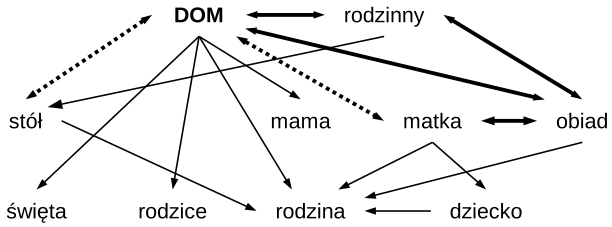


Figure 3. The subnetwork for the sense 'family home – place for the family'

The most important node for the sense 'family home – place for the family' is that for *matka* 'mother', reciprocally linked with that for *dom*. The role played by the mother is indicated by the outgoing links *matka* → *rodzina* 'family', *matka* → *dziecko* 'child', as well as the reciprocal link *matka* ↔ *obiad* 'lunch; dinner' that connects 'place for the family' with the subnetwork *rodzinny* 'family' (attrib.). The role of the mother is additionally augmented by the relatively strong link *dom* → *mama* 'mom' (1.61).

Thus, in the experimental network the node for *matka* plays an immensely significant role: it is the mother who organises the household and the family life, as well as the common meal (*obiad*).[9] It may therefore be concluded that the networked sense 'family home – place for the family' converges with the ethnolinguistic perspective on the Polish *dom* (Bartmiński 2015).

[9] In the framework of traditional Polish lifestyle, *obiad*, the main meal of the day, is usually eaten in the afternoon, so it does not clearly correspond to either lunch or dinner. [translator's note]

The other reciprocal link, complementing the subnetwork 'family home – place for the family', is *dom* ↔ *stół* 'table'. The table is something that organises family space through an outgoing link to the node *rodzina* 'family' (*stół* → *rodzina* 0.81) and a strong incoming link *rodzinny* 'family' (attrib.) → *stół* (2.38).

The last component of the subnetwork that defines 'family home – place for the family' is *rodzina* 'family' (n.), which enters into incoming links only: the very strong *dom* → *rodzina* (12.16) plus the somewhat weaker *matka* 'mother' → *rodzina* (3.72) and *stół* 'table' → *rodzina* (0.81). It was impossible for technical reasons to include in the diagram all the links that would show how the node for *matka* 'mother' constructs the family subnetwork through outgoing links with the nodes *mama* 'mom', *rodzice* 'parents', *ojciec* 'father', *tata* 'dad', and *rodzic* 'parent'.

3.2 Meaning: 'dwelling place'

The skeleton for the subnetwork representing the sense 'dwelling place' are the reciprocal links *dom* ↔ *mieszkanie* 'apartment, flat' (*dom* → *mieszkanie* 15.25; *mieszkanie* → *dom* 12.25), *dom* ↔ *pokój* 'room' (*dom* → *pokój* 1.38; *pokój* → *dom* 2.58); *mieszkanie* 'apartment, flat' ↔ *pokój* 'room' (*mieszkanie* → *pokój* 3.09; *pokój* → *mieszkanie* 8.14), as well as the outgoing links for the node *dom*: *dom* → *mieszkać* 'live, dwell' (0.57), *dom* → *miejsce* 'place' (0.23), *dom* → *własny* 'one's own' (0.8).

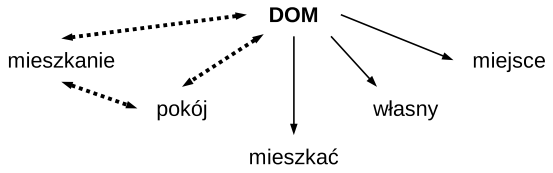


Figure 4. 'House – place for living'

The heart of the sense 'dwelling place' is the node *mieszkanie* 'apartment, flat' together with its links. 'Dwelling place' has properties of a physical object, i.e. it may be owned by the person who lives there, it may be spacious, comfortable, new, small, small but cosy,[10] large, in a block of flats. It may be classified as a luxurious apartment or it may be empty: *dom* ↔ *pusty* 'empty' (*dom* → *pusty* 1.38; *pusty* → *dom* 6.11). The crucial element of 'dwelling place' is the room, which is signaled by reciprocal links: *mieszkanie*

[10] Lit. 'small but one's own' (a fixed collocation in Polish: *ciasne, ale własne*). [translator's note]

'apartment, flat' ↔ *pokój* 'room', *dom* ↔ *pokój* 'room'. The room also has the properties of a physical object listed above.

3.3. Meaning: 'my shelter'

The sense 'my shelter' of *dom* is expressed through the reciprocal link *dom* ↔ *mój* 'my' (*dom* → *mój* 2.18; *mój* → *dom* 10.55) and the outgoing links from the node *dom* to the nodes *bezpieczeństwo* 'security' (1.49), *shelter* (0.46), *azyl* 'refuge' (0.46), *ostoja* 'anchor, lynchpin' (0.23), *twierdza* 'castle' (0.23).

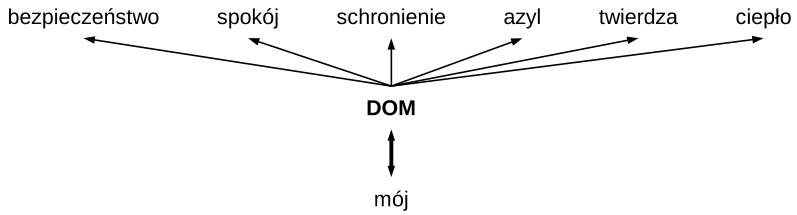


Figure 5. 'My shelter'

The sense 'my shelter' connects through the reciprocal link *pokój* 'room' ↔ *mój* 'my' (*mój* → *pokój* 3.3; *pokój* → *mój* 2.26) with z the node that co-constitutes the skeleton of the meaning of the network that represents the sense 'dwelling place'.

3.4. Meaning: 'building'

This meaning is captured through a subnetwork with no reciprocal links that form its skeletal semantics. Reciprocal links play here the same role as the outgoing and incoming links, i.e. they build the subnetwork. The links of the node *dom* characterise a physical object, i.e. the house's component parts: the chimney, the roof, the walls, windows, the hyperonymous notion of building, the co-hyponym *chata* 'hut', a big or huge size, and the house's location: the countryside, garden, housing estate, or city.

The reciprocal link *duży* 'big'[12] ↔ *dom* (*dom* → *duży* 1.49%; *duży* → *dom* 3.45%) points to a feature of a house that is also characteristic of buildings, cf. the incoming link *budynek* 'building' ← *duży* 'big' (1.13%). The links that have not been included in the diagram, e.g. those with *kościół* 'church', *wysoki* 'tall, high', *cegła* 'brick', *drewno* 'wood', or *drewniany* 'wooden' also corroborate these observations.

[12] The word *duży* 'big' in the experiment had many empty responses. The problem of empty responses is dealt with in Gatkowska (2015a).

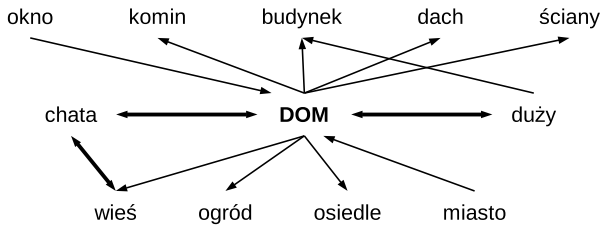


Figure 6. 'Building'

3.5. Meaning: 'institution'

The sense 'institution' is relatively well-represented in the Wortschatz corpus network. No subnetwork represents this meaning in the empirical network, possibly because the relatively fixed collocations *dom dziecka* 'orphanage', *dom kultury* 'community centre', *dom towarowy/handlowy* 'department store', *dom mody* 'fashion house', *dom aukcyjny* 'auction house', *dom maklerski* 'brokerage (firm)', and *dom zakonny* 'monastic house, monastery' function as terms defined legally and outside the linguistic system.

4. *Dom*: network meanings and dictionary meanings

Szymczak's Dictionary of the Polish Language (*SJP* 1978)[13] provides six meanings of *dom*, arranged (as one can presume) in the order of importance:
1. 'a building designed for living purposes, for companies, institutions, etc.';
2. 'apartment, permanent living place';
3. 'family, household';
4. 'the totality of family and domestic matters; homestead';
5. 'clan, family, dynasty';
6. 'a state, social, trading etc. institution, usually located in a separate apartment or building; the building itself'.

Our analysis of the subnetworks that represent the meanings of *dom* indicates that the experimental network constructs the meaning hierarchy differently, e.g. the sense 'building', which occupies the first position in the dictionary, comes as last but one. One can also clearly see differences in the set of meanings being identified. The common meanings are undoubtedly 1. 'building' and 2. 'apartment, flat'. Sense 3. 'family, household' is organised in the network via the *matka* 'mother' subnetwork and has a different structure

[13] A similar account is found in the Internet Dictionary of Polish, *SJP PWN* (http://sjp.pwn.pl/sjp/; accessed 15 March, 2016), which also distinguishes the sense 'place of one's origin'.

than that suggested by the dictionary definition. Sense 6. 'institution' is not represented as a distinct subnetwork at all. There are also significant differences in the meaning sets: the network lacks dictionary senses 4. 'the totality of family and domestic matters; homestead' or 5. 'clan, family, dynasty', whereas the dictionary does not record the sense 'my shelter', rather significant in the experimental network structure. We confine our observations here to merely pinpointing the differences, without trying to explain them away, since the latter task would have definitely taken us beyond the scope of the present study.

5. Synopsis

The article presents the semantics of the Polish *dom* 'house/home' in empirical networks. Each of the two networks discussed had been created differently but they both contain a similar hierarchy of the meanings of *dom*: 'place for the family', 'place for living', 'building', 'institution'. The experimentally constructed network contributes a new quality to the picture, as it contains subnetworks that explain each of the word's meanings. A comparison of the networked description of meanings with lexicographic definitions shows that the experimental network can reveal meanings absent from dictionaries. Empirical lexical networks can thus be helpful in lexicography.

An analysis of meanings based on the structural properties of the experimental network, however, does not provide a complete description of networked semantics because the stimulus-response dependence is semantic in nature (Clark 1970), e.g. the link *dom* → *dach* 'roof' can only be explained as one that obtains between the whole and one of its parts (metonymy). Needless to say, an analysis of all semantic relationships between network nodes calls for a distinct and a much more detailed treatment (cf. Gatkowska 2015b).

References

Bartmiński, Jerzy. 2015. DOM – koncept uniwersalny i specyficzny kulturowo. In: Jerzy Bartmiński, Iwona Bielińska-Gardziel, and Beata Żywicka (eds.) *Leksykon aksjologiczny Słowian i ich sąsiadów*, vol. 1, *Dom*. 15–33. Lublin: Wydawnictwo UMCS.

Biemann, Chris, Gerhard Heyer, Uwe Quasthoff, and Matthias Richter. 2007. The Leipzig Corpora Collection – Monolingual corpora of standard size. In: *Proceedings of Corpus Linguistics 2007*. Birmingham, UK. (http://wortschatz.uni-

leipzig.de/~cbiemann/pub/2007/BiemannHeyerQuasthoffRichterCorpLing07.pdf; accessed Dec 15, 2015)

Clark, Herbert H. 1970. Word associations and linguistic theory. In: John Lyons (ed.) *New Horizons in Linguistics*. 271–286. Harmondsworth, Middlesex: Penguin Books.

De Deyne, Simon and Ger Storms. 2008. Word associations: Network and semantic properties. *Behavior Research Methods* 40 (1): 213–231.

Gatkowska, Izabela. 2014. Word associations as a linguistic data. In: Piotr Chruszczewski, John R. Rickford, Katarzyna Buczek, Aleksandra Knapik, and Jacek Mianowski (eds.) *Languages in Contact 2012*, vol. 1. 79–92. Wrocław: Wydawnictwo WSF we Wrocławiu.

Gatkowska, Izabela. 2015a. Empty answers in an experiment of free word association. *Academic Journal of Modern Philology* 4: 41–49.

Gatkowska, Izabela. 2015b. Empiryczna sieć powiązań leksykalnych. *Polonica* 35: 155–178.

Gawarkiewicz, Roman. 2008. Dom w językowym obrazie świata młodzieży polskiej i rosyjskiej. In: Roman Gawarkiewicz, Izabela Pietrzyk, and Barbara Rodziewicz (eds.) *Polski słownik asocjacyjny z suplementem*. 235– 246. Szczecin: Print Group.

Gawarkiewicz, Roman, Izabela Pietrzyk, and Barbara Rodziewicz. 2008. *Polski słownik asocjacyjny z suplementem*. Szczecin: Print Group.

Kent, Grace H. and Aaron J. Rosanoff. 1910. A study of association in insanity. *American Journal of Insanity* 67 (1–2): 317–390.

Kiss, George R., Christine Armstrong, Robert Milroy, and James Piper. 1973. An associative thesaurus of English and its computer analysis. In: A. J. Aitken, Richard W. Bailey, and N. Hamilton-Smith (eds.) *The Computer and Literary Studies*. 153–165. Edinburgh: Edinburgh University Press.

Kurcz, Ida. 1967. Polskie normy powszechności skojarzeń swobodnych na 100 słów z listy Kent-Rosanoffa. *Studia Psychologiczne* 8: 122–255.

Minsky, Marvin. 1975. A framework for representing knowledge. In: Patrick H. Winston (red.) *The Psychology of Computer Vision*. 211–277. New York: McGraw-Hill.

Palermo, David S. and James J. Jenkins. 1964. *Word Association Norms: Grade School through College*. Minneapolis: University of Minnesota Press.

Postman, Leo Joseph and Geiffrey Keppel. 1970. *Norms of Word Association*. New York: Academic Press.

Rosenzweig, Mark R. 1957. Etudes sur l'association des mots. *L'Année Psychologique* 57: 23–32.

Rosenzweig, Mark R. 1961. Comparisons among word-association responses in English, French, German, and Italian. *American Journal of Psychology* 74: 347– 360. doi: http://dx.doi.org/10.2307/1419741

Russell, Wallace A. and James J. Jenkins. 1954. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff word association test. Technical Report No. 11, University of Minnesota, ONR, Contract No. N8 onr-66216.

Russell, Wallace A. and O. R. Meseck. 1959. Der Einfluss der Association auf das Erinnern von Worten in der deutschen, franzosischen and englischen Sprache. *Zeitschrift für Experimentelle und Angewandte Psychologie* 6: 191–211.

Schank, Roger C. 1975. *Conceptual Information Processing*. Amsterdam: North-Holland.

Schulte im Walde, Sabine, Susanne Borgwaldt and Ronny Jauch. 2012. Association norms of German noun compounds. In: Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Ugur Dogan, Bente Maegaard, Joseph Mariani, Jan Odijk,

Stelios Piperidis (eds.) *Proceedings of the 8th International Conference on Language Resources and Evaluation, LREC 2012, Istanbul, Turkey, May 23-25, 2012*. 632-639. European Language Resources Association.
Schellenberg, P. E. 1930. A Group Free-association Test for College Students. Unpublished doctoral dissertation. University of Minnesota.
SJP. 1978. *Słownik języka polskiego*. Ed.-in-Chief: Mieczysław Szymczak. Warszawa: PWN.

Appendix
Loose word association test as a tool in lexical network construction

1. The association test: a brief history

The association test is a method of accessing lexical dependencies stored in the human mind. The procedure is simple: the subject is asked to supply the first word that comes to mind upon understanding the stimulus word provided by the experimenter. The method was first used for scientific by two American psychiatrists, Grace Kent and Aaron Rosanoff (1910), who envisaged and implemented the procedure for 100 words with 1,000 healthy subjects. The goal of the experiment was to construct a diagnostic model of association norms as a frame of reference for comparing association sets elicited from persons with mental disorders.

Kent and Rosanoff's procedure was replicated with a group of American students by Schellenberg (1930; referenced in Postman and Keppel 1970). The next time it was replicated was over twenty years later by Wallace Russell and James Jenkins (1954), also with American college students. The replicated studies yielded results that were surprisingly convergent with those of the original test, which was interpreted as meaning that lexical connections are stable and independent from the actual make-up of the test group. Therefore, in subsequent studies the decision was made to reduce the number of subjects and increase that of the stimulus words, e.g. to 500 subjects and 200 words in Palermo and Jenkins (1964), or to 100 subjects and 8,400 words in Kiss et al. (1973), the latter test being conducted in Great Britain. The Kiss et al.'s team also modified the procedure, being the first to run it in phases: the initial set of stimulus words contained just over 2,000 words, the responses to which were then used as stimuli in the second phase. On the basis of the results obtained, *The Edinburgh Associative Thesaurus* (EAT) was constructed, the first project of this kind in the history of the lexical associations experimental network research.

In the second half of the 20th c. a new chapter was opened in this field of research by translating the Kent–Rosanoff list into several languages. Associative studies were conducted for French (Rosenzweig 1957), German (Russell and Meseck 1959), Italian (Rosenzweig 1961), or Polish (Kurcz 1967, an experiment with 1,000 Warsaw-based students).

Two trends in lexical association test research can clearly be identified since EAT. The first one is studying psycholinguistic mechanisms in a relatively small groups of subjects and with relatively limited sets of stimuli. This trend is mainly present in psychology and as such will not be the focus of our attention. The second trend is studying the structure of the lexicon and lexical combinability with relatively many subjects and with the use of large sets of stimuli, e.g. De Deyne and Storms (2008). Not infrequently, studies of this kind do not strictly follow the rigour of the psychological experiment: they are conducted online and the subjects are asked to provide more than one word that they associate with the stimulus (e.g. Schulte im Valde, Borgwaldt and Jauch 2012). In the Polish context, the only large-scale study (500 subjects and 110 words) was carried out by Gawarkiewicz, Pietrzyk and Rodziewicz (2008) in Szczecin. The goal of that study was to construct an associative dictionary as an aid in Polish-Russian contrastive research; the procedure involved one associative cycle. As a result, the dictionary is similar to the association norms (lists) obtained by Kurcz: the two only differ with regard to the method of result presentation. Kurcz's norms include the stimulus and the set of responses obtained, each being accompanied by the number of subjects that provided the response. In the associative dictionary there are two kinds of lists. The first type, similarly to Kurcz's, provides the stimulus and the set of responses with the number of the subjects. The second type gives a response and the set of stimuli words that had invoked that response – importantly, the number of subjects that linked a given stimulus with the response is provided for the stimulus (rather than for the response). This kind of list does not appear in Kurcz, since the author does not consider them to be association norms. Looking at the associative dictionary from the network point of view one can say that lists of the first type are stimulus-based (and concern outgoing links), whereas lists of the second type are response-based (and concern incoming links). The lexical network nodes constructed around the stimulus are far richer structures than lists, even if one omits the information in the lexical node paths (a basic picture of the content of the paths can be seen in the diagram for the *dom* node). The differences between the associative list and the lexical network node are clearly shown through the specification of the direct links of the stimulus words. The associative list of the stimulus in the associative

dictionary only contains outgoing links, whereas the stimulus node in the network includes both outgoing and incoming links, which renders the node a semantically richer structure. Table App. 1 shows this with the aid of the strongest links for the adjective *biały* 'white'. For the sake of comparison, the linkage strength in DCL JU is omitted and only the number of responses is given.

Table App. 1. Strongest links for *biały* 'white' in the associative dictionary and the DCL JU studies

Associative dictionary		DCL JU lexical network			
		outgoing links		incoming links	
czarny 'black'	128	*śnieg* 'snow'	181	137	orzeł 'eagle'[i]
kolor 'colour'	78	*kolor* 'colour'	111	131	*czarny* 'black'
śnieg 'snow'	61	*czarny* 'black'	79	66	*fartuch* 'apron'
czystość 'cleanliness'	28	*kruk*, lit. 'raven'[ii]	43	47	*ser* 'cheese'[iii]
czysty 'clean'	26	*dom* 'house'[iv]	34	39	*cukier* 'sugar'
niewinność 'innocence'	18	*miś* 'teddy bear'[v]	28	30	*mleko* 'milk'
bałwan 'snowman'	7	*obrus* 'tablecloth'	26	21	*dym* 'smoke'
anioł 'angel'	6	*koń* 'horse'	17	20	*bielinek* 'pierid'
kruk, lit. 'raven'[ii]	6	*ser* 'cheese'[iii]	17	13	*jasny* 'bright'
mleko 'milk'	6	*kieł* 'fang'[vi]	15	9	*baran* 'ram'
orzeł 'eagle'[i]	6	*orzeł* 'eagle'[i]	15	5	*owca* 'sheep'
dobry 'good'	5	*dzień* 'day'[vii]	13	4	*doktor* 'doctor'
gołąb 'dove'	5	*papier* 'paper'	11	4	*lekarz* 'doctor'
flaga 'flag'[viii]	4	*kot* 'cat'	10	4	*motyl* 'butterfly'

i *Biały orzeł* or officially *Orzeł Biały* is the Polish national emblem
ii an idiomatic fixed collocation: *biały kruk* 'a rare book; rarity'
iii a fixed collocation: *biały ser* 'cottage cheese'
iv often *Biały Dom* 'the White House'
v *Biały miś* 'A white teddy bear' is a popular pop song
vi cf. Jack London's popular novel *White Fang*
vii a fixed collocation *w biały dzień* 'in broad daylight'
viii *Biała flaga* 'White flag' is a popular rock song

A comparison of the outgoing links obtained in both studies reveals the already mentioned replicability of the associative test results, the differences in the order of the links resulting from the way the test is administered (cf. below), as well as from the unequal number of subjects. The incoming links of the network node, in turn, combine with the outgoing ones into reciprocal links, e.g. *biały* 'white' → *czarny* 'black' and *biały* ← *czarny*, as well as enriching the stimulus's semantic profile, e.g. *biały* 'white' ← *fartuch* 'apron', *mleko* 'milk', *cukier* sugar', etc.

2. The DCL JU cyclical experiment

a. Selection of stimuli

Because the goal of the DCL JU study was to construct an experimental lexical network, it had been designed as a cyclical procedure. In the first cycle 63 nouns were used as stimulus words (the Polish version of a portion of the Kent–Rosanoff list, cf. Kurcz 1967) – these are the so called primary stimuli. In the second cycle the stimuli were the strongest associations with the primary stimuli (five associations with each primary stimulus): these are the 259 secondary stimuli. The reduced number of secondary stimuli (63 x 5 amounts to 315, vs. the actually used 259) results from the fact that if a word appeared in the top five most frequent responses to a few primary stimuli, that word appeared in the set of secondary stimuli only once. 322 stimulus words were used in the first and second cycle of the study in total. The cyclicity of the experiment is a novelty in research on the Polish lexicon and distinguishes our approach from that represented by Kurcz or Gawarkiewicz.

b. The subjects

The experiment, conducted in the years 2011-2014, involved 900 students of the Jagiellonian University and the University of Science and Technology in Kraków. It was carried out anonymously, with each subject only providing their age and sex. Among the subjects there were no students of philology or psychology because the knowledge and expertise that they would have obtained in the course of their studies might have influenced the test results.

c. The procedure

The experiment was conducted under controlled conditions. A specially designed computer system was used, another novelty in comparison with previous research, where paper questionnaires were used (Kurcz 1967; Gawarkiewicz 2008). Each participant worked with a computer, responding to stimulus words that were successively displayed on the screen. The task was to type in the first word that came to mind upon reading and actually understanding the stimulus word. The subject was only given five seconds to start typing in the response – after this time the system would automatically consider it to be nil and would move on to the next stimulus. The subjects could not go back to the words for which the response had not been provided, nor could they change the responses already given, which was possible when working with paper questionnaires.

In short, one can say that the computer-aided procedure helps one obtain spontaneous and unpremeditated responses, which also provides valuable

psychological data. For example, registering the response time facilitates a construction of subject profiles.

d. The experimental lexical network characterised

The experiment has produced a network of 50,849 links and 11,224 lexical nodes, including 322 full nodes. The remaining nodes are of the reduced type, i.e. they only involve incoming links. The network also contains 1,181 word (node) pairs of the reciprocal-link type.

The set of nodes in the DCL JU lexical network includes:
– 7,757 nouns (69.1% of the total);
– 2,459 adjectives (21.9%);
– 744 verbs (6.6%);
– 264 items belonging to other categories (2.4%).

The list reveals substantial similarity between the experimentally constructed network for Polish and a similar network for Dutch (De Deyne and Storms 2008), where the nodes are:
– nouns (72%);
– adjectives (18%);
– verbs (9%);
– other (1%).

The structure of a network lexical node partly depends on the number of subjects involved in the experiment. A stimulus in the DCL JU network (900 subjects) has on average 150 outgoing links, whereas in EAT (100 subjects) it only has 50 direct links. As a result, the former network has a richer structure, which allows one to find in a lexical node the paths that explain distant links, e.g. *baranina* 'mutton' – *sweter* 'pullover'. An EAT node has a less complex structure, which frequently does not allow one to explain distant links.

translated by Adam Głaz