

- **Dane słownikowe** – w jaki sposób jak najmniejszym wysiłkiem uzyskać wszystkie formy słów, powiązania form odmienionych z formami hasłowymi, powiązania słów z dodatkową informacją.
- **Analiza morfologiczna** – jak uzyskać cechy składniowe słowa, formę podstawową na podstawie formy odmienionej, podział słowa na części składowe. **Synteza morfologiczna** – jak uzyskać wszystkie formy odmienione z formy podstawowej i dodatkowych informacji.
- **Struktury danych** – w jaki sposób tworzyć i przechowywać słownik, tak aby zajmował jak najmniej miejsca i gwarantował szybki dostęp do informacji.

# Morfemy (1/2)

- **Morfem** jest (w uproszczeniu) elementem składowym słowa. Wyróżniamy morfemy swobodne i morfemy związane.
- **Morfem swobodny** może wystąpić samodzielnie, nie w połączeniu z innymi morfemami. **Morfem związany** nie ma tej właściwości.
- **Rdzeń** to podstawowy morfem wyrazu, do którego mogą być dołączane morfemy afiksalne (afiksy), zarówno fleksyjne, jak i słowotwórcze. Rdzeń jest **morfemem leksykalnym**. Przykład: do**dat**kowy.
- **Temat** wyrazu to ta część wyrazu, która pozostaje po usunięciu afiksów fleksyjnych. Jest nośnikiem znaczenia wyrazu. Nie musi być pojedynczym morfemem. Przykład: **dodat**kowy.

- **Afiksy** to morfemy związane niosące funkcje gramatyczne. Dzielą się na **przedrostki**, **przyrostki** i **wrostki**, czasami wyróżnia się też postfiksy (**zarostki?**). Istnieją też morfemy nieciągłe, dołączane jednocześnie na początku i końcu wyrazu (**obrostki?**).
- **Morfem zerowy** niesie jedynie informację gramatyczną, natomiast nie posiada reprezentacji tekstowej. Np. mianownik liczby pojedynczej wyrazu **stoń** może być przeanalizowany jako **stoń+∅**, czyli temat **stoń** i morfem zerowy.

# Morfologiczne typy języków

- Języki **izolujące** nie posiadają form związanych, które mogą być dołączane do słowa.
- W językach **aglutynacyjnych** poszczególne formy związane występują i są układane w słowie jak paciorki różańca; jedna forma związana reprezentuje pojedynczą cechę.
- W językach **fleksyjnych** wiele różnych cech jest reprezentowanych przez pojedynczą formę związaną lub ściśle powiązane ze sobą formy związane.
- W językach **polisyntetycznych** pewne elementy, które w innych językach pojawiają się osobno, jak np. dopełnienie, są wyrażane za pomocą środków morfologicznych.

- W **morfologii fleksyjnej** opisujemy różne formy tego samego samego słowa – **leksemu**. Zbiór wszystkich form odmienionych opisuje **wzór odmiany** lub **paradygmat** danego słowa. Formy odmienione reprezentują tę samą część mowy co wyraz hasłowy, mają to samo znaczenie, ale dostosowują wyraz do pełnienia innych funkcji składniowych.
- **Morfologia derywacyjna** to inaczej **słowotwórstwo**. Na podstawie jednego leksemu otrzymujemy inny, najczęściej będący inną częścią mowy, o innym znaczeniu.

- Fleksja:

**droga** → drogi, drodze, drogę, drogi, dróg...

**drogi** → drogiego... droga... drodze...

**pies** → psa, psu, psie, psy...

**pieszy** → pieszego, pieszemu, pieszym, piesi...

**kochać** → kocham, kochasz... kochałem... kochałabyś...

- Słowotwórstwo:

**pies** → piesek, pieseczek, psi, pieski...

**kot** → kotek, koteczek, kocię, koci, kocić się...

**kochać** → kochanie, kochanek, kochanka, kochliwy...

Najprostszą i jednocześnie najpowszechniejszą formą łączenia morfemów jest **sklejanie** (ang. *concatenation*) — tekstowe przyłączanie morfemu (przyrostka) na końcu lub (przedrostka) na początku wyrazu. Rzadszą formą jest dołączanie morfemu w środku wyrazu. Operacja ta podlega różnym ograniczeniom i bywa traktowana jako forma dołączania przedrostka przy pominięciu pewnych początkowych elementów wyrazu, jak np. pierwszej spółgłoski, pierwszej sylaby, pierwszej sylaby nieakcentowanej itp.

W językach semickich temat czasowników jest wzorcem składającym się z rdzenia (na ogół trzech spółgłosek), wzorca samogłosek (informującego o stronie i aspekcie) i wzorca słowotwórczego (określającego klasę).

Np. *ktb* oznacza *pisać*. W stronie czynnej wzór samogłosek zawiera **A**, więc słowo to w stronie czynnej dla wzorca słowotwórczego **CVCVC** (forma podstawowa) brzmi *kAtAb*, zaś w stronie biernej wzór samogłosek ma postać **UI**, a samo słowo – *kUtIb*. Wzorzec słowotwórczy **CVVCVC** prowadzi do słowa *korespondować*: dla strony czynnej – *kAAAtAb*, biernej – *kUUtIb*.



## Łączenie morfemów – ograniczenia (1/2)

Nie każdy afiks łączy się z każdym słowem w dowolnym miejscu. Ograniczenia mogą dotyczyć części mowy, przypadku i innych cech gramatycznych, a także cech wymowy i znaczenia. Do opisu ograniczeń można użyć różnych sposobów, np. klas kontynuacji. Najprostsze wydaje się użycie unifikacji. Przykład (przysłówki w stopniu równym) w programie mmorph:

```
adv_pos : adv[deg=pos advs=$adv form=surface]
          ← a[deg=pos form=stem advs=$adv par_a!=no]
            advsuf[deg=pos advs=$adv]
```

## Łączenie morfemów – ograniczenia (1/2)

Nie każdy afiks łączy się z każdym słowem w dowolnym miejscu. Ograniczenia mogą dotyczyć części mowy, przypadku i innych cech gramatycznych, a także cech wymowy i znaczenia. Do opisu ograniczeń można użyć różnych sposobów, np. klas kontynuacji. Najprostsze wydaje się nazwa reguły akcji. Przykład (przysłówki w stopniu równym) w programie morph:

```
adv_pos : adv[deg=pos advs=przysłówek]
← a[deg=pos form=stem advs=przysłówek par_a!=no]
adv_suf[deg=pos advs=$adv]
```

przysłówek jest tworzony...

...z przymiotnika...

...i końcówki przysłówka

## Łączenie morfemów – ograniczenia (1/2)

Nie każdy afiks łączy się z każdym słowem w dowolnym miejscu. Ograniczenia mogą dotyczyć części mowy, przypadku i innych cech gramatycznych, a także cech wymowy i znaczenia. Do opisu ograniczeń można użyć różnych sposobów, np. klas kontynuacji. Najprostsze wydaje się użycie **stopień...** i **...i wzór odmiany zgodne** kontynuacji. Przykład (przysłówki w stopniu równym) w programie mmorph:

```
adv_pos : adv[deg=pos advs=$adv form=surface]
          ← a[deg=pos form=stem advs=$adv par_a!=no]
            advsuf[deg=pos advs=$adv]
```

## Łączenie morfemów – ograniczenia (2/2)

Końcówki do reguły:

adv.pos1: "o" advsuf[deg=pos advs=o]

adv.pos2: "&prim;e" advsuf[deg=pos advs=e]

Słownik (fragment):

a[deg=pos form=stem par\_a=y dega=no advs=o]

"bos" = "bosy"

"bezosobow" = "bezosobowy"

"burzow" = "burzowy"

"brodat" = "brodaty"

...

# Zmiany w morfemach wynikające z łączenia

	pies	
	lp	lm
M	pies	psy
D	psa	psów
C	psu	psom
B	psa	psy
W	psie	psy
N	psem	psami
Mc	psie	psach

Co jest tematem w leksemie **pies**?  
**pies** czy **ps**?  
Jeden z nich, oba, czy jeszcze coś innego?

	pies	
	lp	lm
M	pies+∅	ps+y
D	ps+a	ps+ów
C	ps+u	ps+om
B	ps+a	ps+y
W	ps+ie	ps+y
N	ps+em	ps+ami
Mc	ps+ie	ps+ach

Mamy dwa tematy: podstawowy **pies** i poboczny **ps** (albo odwrotnie).

Rozwiązanie to wydaje się proste i jest często stosowane (np. nawet w INTEX/Unitex). Wymaga jednak podawania zestawu tematów dla każdego słowa, w którym zachodzą takie oboczności i jest podatne na błędy.

	pies	
	lp	lm
M	p's+∅	p's+y
D	p's+a	p's+ów
C	p's+u	p's+om
B	p's+a	p's+y
W	p's+'e	p's+y
N	p's+em	p's+ami
Mc	p's+'e	p's+ach

Temat leksemu **pies** to **p's**, czyli miękkie **p'** i **s**. Poza M lp po miękkim **p** występuje **s**, więc miękkość **p** nie przejawia się w żaden widoczny sposób. W M lp dołączamy morfem zerowy. Ponieważ **ps** jest trudno wymówić, do środka wstawiane jest **e**. Teraz miękkie **p'** może się ujawnić: mamy **pie**.

## Spółgłoskowe:

**m:m', b:b', p:p', v:v'**: łamać – łamię, krok – krokiem, pies – psa

**t:ć, c:c', d:dź, d:dż, s:ś, sz:ś, z:ż, ż:ż, n:ń**: wożę – wozisz, noszę – nosisz

**r:rz**: karać – karzę, brać – bierzesz

**k:cz, g:ż**: piekę – pieczesz, mogę – możesz

## Samogłoskowe:

**e:o, e:a, o:ó, ę:ą**: bierzemy – biorę, księga – książkę

Utrata lub wstawienie samogłoski: samogłoska – samogłosek

Wymiany głoskowe mogą występować wielokrotnie w tym samym słowie:

gwiazda – gwieździe, brać – biorę



W języku niemieckim występują **przegłosy** (niem. *umlaut*) oznaczane w piśmie za pomocą specjalnego znaku diakrytycznego. Samogłoska bez przegłosu w pierwszej sylabie rzeczownika może zmieniać się na samogłoskę z przegłosem w liczbie mnogiej: **Land** – **Länder**, **Bruder** – **Brüder**.

W wielu językach, w tym semickich, ąłtajskich i ugrofińskich, występuje zjawisko **harmonii samogłosek**. Polega na zgodności wybranych wyróżniających cech fonologicznych samogłosek afiksów z cechami fonologicznymi ostatniej samogłoski rdzenia (np. otwartość samogłosek).